



**HAL**  
open science

# Mean-field limits for Consensus-Based Optimization and Sampling

Nicolai Jurek Gerber, Franca Hoffmann, Urbain Vaes

► **To cite this version:**

Nicolai Jurek Gerber, Franca Hoffmann, Urbain Vaes. Mean-field limits for Consensus-Based Optimization and Sampling. 2024. hal-04346646v2

**HAL Id: hal-04346646**

**<https://hal.science/hal-04346646v2>**

Preprint submitted on 30 Aug 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - ShareAlike 4.0 International License

# Mean-field limits for Consensus-Based Optimization and Sampling

Nicolai Jurek Gerber<sup>a,1</sup>, Franca Hoffmann<sup>b,2</sup>, and Urbain Vaes<sup>c,3,4</sup>

<sup>a</sup>[nicolai.gerber@uni-bonn.de](mailto:nicolai.gerber@uni-bonn.de), <sup>b</sup>[franca.hoffmann@caltech.edu](mailto:franca.hoffmann@caltech.edu), <sup>c</sup>[urbain.vaes@inria.fr](mailto:urbain.vaes@inria.fr)

<sup>1</sup>Hausdorff Center for Mathematics, Rheinische Friedrich-Wilhelms-Universität, Bonn, Germany

<sup>2</sup>Department of Computing and Mathematical Sciences, Caltech, USA

<sup>3</sup>MATERIALS project-team, Inria Paris

<sup>4</sup>CERMICS, École des Ponts, France

16th August 2024

## Abstract

For algorithms based on interacting particle systems that admit a mean-field description, convergence analysis is often more accessible at the mean-field level. In order to transpose convergence results obtained at the mean-field level to the finite ensemble size setting, it is desirable to show that the particle dynamics converge in an appropriate sense to the corresponding mean-field dynamics. In this paper, we prove quantitative mean-field limit results for two related interacting particle systems: Consensus-Based Optimization and Consensus-Based Sampling. Our approach requires a generalization of Sznitman’s classical argument: in order to circumvent issues related to the lack of global Lipschitz continuity of the coefficients, we discard an event of small probability, the contribution of which is controlled using moment estimates for the particle systems. In addition, we present new results on the well-posedness of the particle systems and their mean-field limit, and provide novel stability estimates for the weighted mean and the weighted covariance.

**Keywords.** Mean-field limits, Interacting particle systems, Consensus-Based Optimization, Consensus-Based Sampling, Coupling methods, Wasserstein stability estimates.

**AMS subject classification.** 35Q93, 65C35, 70F45, 35K55.

## 1 Introduction

This work focuses on the rigorous derivation of quantitative mean-field limits for recently proposed algorithms known as Consensus-Based Optimization (CBO) and Consensus-Based Sampling (CBS), as well as their variants. The former, initially proposed in [46], is an algorithm for global optimization tasks, while the latter, proposed in [12], can be employed for both sampling and optimization purposes. Both methods are based on an interacting particle system with only locally, not globally, Lipschitz coefficients and interaction kernels. This introductory section is organized as follows: we first introduce Consensus-Based Optimization (Subsection 1.1) and Consensus-Based Sampling (Subsection 1.2), giving pointers to the relevant literature. In Subsection 1.3 we present a brief review of the literature on the use of coupling methods to prove propagation of chaos for models with non-globally Lipschitz coefficients. Finally, we summarize the main contributions of this work in Subsection 1.4, and introduce the notation used throughout the paper in Subsection 1.5.

Since Consensus-Based Sampling can be viewed as a variant on Consensus-Based Optimization, we focus predominantly on the former and then briefly discuss the latter.

## 1.1 Consensus-Based Optimization (CBO)

Given an objective function  $f: \mathbf{R}^d \rightarrow \mathbf{R}$ , we consider the following version of Consensus-Based Optimization (CBO):

$$dX_t^{(j)} = - \left( X_t^{(j)} - \mathcal{M}_\beta(\mu_t^J) \right) dt + S \left( X_t^{(j)} - \mathcal{M}_\beta(\mu_t^J) \right) dW_t^{(j)}, \quad t \geq 0, \quad j \in \llbracket 1, J \rrbracket, \quad (1.1)$$

where  $J > 1$  is a natural number,  $\llbracket 1, J \rrbracket := \{1, \dots, J\}$  and  $\mu_t^J := \frac{1}{J} \sum_{i=1}^J \delta_{X_t^{(i)}}$  denotes the empirical measure associated with the particle system. For a probability measure  $\mu \in \mathcal{P}(\mathbf{R}^d)$ , the notation  $\mathcal{M}_\beta(\mu)$  denotes the mean of  $\mu$  after reweighting by  $e^{-\beta f}$ . Its precise definition, together with its expression when  $\mu = \mu_t^J$ , is given by

$$\mathcal{M}_\beta(\mu) := \frac{\int_{\mathbf{R}^d} x e^{-\beta f(x)} \mu(dx)}{\int_{\mathbf{R}^d} e^{-\beta f(x)} \mu(dx)}, \quad \mathcal{M}_\beta(\mu_t^J) := \frac{\sum_{j=1}^J X_t^{(j)} e^{-\beta f(X_t^{(j)})}}{\sum_{k=1}^J e^{-\beta f(X_t^{(k)})}}. \quad (1.2)$$

The weighted mean  $\mathcal{M}_\beta(\mu_t^J)$  may be viewed as an approximation of the particle position  $X_t^{(j)}$  such that  $f(X_t^{(j)})$  is minimal. In (1.1), the processes  $(W_t^{(j)})_{j \in \llbracket 1, J \rrbracket}$  are independent  $\mathbf{R}^d$ -valued Brownian motions and  $S: \mathbf{R}^d \rightarrow \mathbf{R}^{d \times d}$  is a globally Lipschitz continuous function. The two diffusion types commonly studied in the literature are isotropic diffusion [46, 11, 24], with  $S(u) = \sqrt{2\theta}|u| \mathbf{I}_d$ , and anisotropic or component-wise [14, 25] diffusion with  $S(u) = \sqrt{2\theta} \text{diag}(u)$ . Here  $\mathbf{I}_d \in \mathbf{R}^{d \times d}$  is the identity matrix and  $\text{diag}: \mathbf{R}^d \rightarrow \mathbf{R}^{d \times d}$  the operator mapping a vector onto a diagonal matrix with the vector as its diagonal. It was shown in [14] that CBO with anisotropic diffusion is well-suited for high-dimensional global optimization problems, for example those arising in machine learning. Commonly a constant coefficient is included in front of the drift term in (1.1), but since this coefficient can be removed by rescaling time upon redefining the diffusion parameter  $\theta$ , we choose to omit it.

By design, CBO allows passage to the corresponding mean-field description. Indeed, since the particles in (1.1) evolve towards the weighted mean (1.2) instead of the current best particle, as is the case for Particle Swarm Optimization (PSO) [38], it is possible to formally write down the corresponding mean-field equation:

$$\begin{cases} d\bar{X}_t = - (\bar{X}_t - \mathcal{M}_\beta(\bar{\rho}_t)) dt + S(\bar{X}_t - \mathcal{M}_\beta(\bar{\rho}_t)) dW_t, \\ \bar{\rho}_t = \text{Law}(\bar{X}_t), \end{cases} \quad (1.3)$$

where the weighted mean  $\mathcal{M}_\beta(\mu)$  for general measures  $\mu \in \mathcal{P}(\mathbf{R}^d)$  can be interpreted as a smooth approximation of the 'current best' that appears in PSO [30]. The connection between the particle system (1.1) and the dynamics (1.3), which is of McKean–Vlasov type and comparatively simpler to analyze mathematically, stimulated theoretical research using tools from stochastic analysis and partial differential equations [35, 11, 24]. In particular, consensus formation and convergence of the CBO mean-field dynamics (1.3) to a consensus point as  $t \rightarrow \infty$  has been proven in [11]. The distance between the consensus point and the true global minimizer can be made arbitrarily small by choosing a large exponential parameter  $\beta$ . The proof in [11] covers also non-convex cost functions  $f$ .

The convergence, in an appropriate sense, of the empirical measure associated with the CBO finite-size particle system (1.1) to the solution of the mean-field equation (1.3) in the large particle limit  $J \rightarrow \infty$  was first proved rigorously in a recent paper [35] using a compactness argument, as we discuss further in Subsection 1.3. However, the method of proof employed in [35] does not provide an explicit rate of convergence with respect to the number of particles. Obtaining such an explicit rate of convergence is highly desirable, as it would enable to transfer convergence results concerning the mean-field equation (1.3) to the finite ensemble size setting (1.1) by means of a triangle inequality argument, modulo appropriate interfacing

of the metrics. This is the approach proposed in [24, Theorem 13], where the authors obtain a partial mean-field result in a set of high probability, enabling them to prove convergence with high probability to the global minimizer in the finite ensemble size setting. Here, we provide a stronger version of the mean-field limit result given in [24, Proposition 15], extending to more general diffusion terms and providing scope for global convergence results in a stronger metric than convergence in probability.

A number of extensions of CBO have been proposed in the literature. As mentioned previously, using an anisotropic diffusion term results in a scheme which is suitable for high-dimensional state spaces, and has been applied successfully in the training of artificial neural networks [14]. Using componentwise geometric Brownian motion in this way allows to prove consensus formation directly at the particle level, which was achieved in [31, 32, 42]. (On this subject, we point out that the results in [32] on the proximity of the consensus point to the global minimizer concern only the best case scenario; a more general convergence result in probability or a different metric is still lacking.) We also mention adaptations with mini-batching [36], memory effects [49], and Lévy noise [37], as well as extensions to multi-objective optimization [41, 8] and global optimization with constraints [22, 23, 13, 9, 33, 4]. Moreover, a variant tailored for objective functions with several minima was proposed in [10], and results establishing a link between CBO and PSO in the so-called small inertia limit were obtained in [30, 18]. Many of the recent advances concerning the mathematical analysis of CBO and its extensions were recently overviewed in [48]. Furthermore, a free software implementation of CBO and some of its extensions was recently published [5].

## 1.2 Consensus-Based Sampling (CBS)

Consensus-Based Sampling is a sampling method inspired by the Ensemble Kalman Sampler (EKS) [27] and by consensus-based approaches such as CBO. It is based on an interacting particle system with a noise term preconditioned by the square root of the empirical covariance. The method in the finite ensemble size and continuous time setting reads

$$dX_t^{(j)} = -\left(X_t^{(j)} - \mathcal{M}_\beta(\mu_t^J)\right) dt + \sqrt{2\lambda^{-1}\mathcal{C}_\beta(\mu_t^J)} dW_t^{(j)}, \quad j \in \llbracket 1, J \rrbracket, \quad (1.4)$$

with parameters  $\beta, \lambda > 0$ . Here, for a probability measure  $\mu \in \mathcal{P}(\mathbf{R}^d)$ , the notation  $\mathcal{C}_\beta(\mu)$  denotes the covariance of  $\mu$  after reweighting by  $e^{-\beta f}$ ; see (1.9) in Subsection 1.5 for the precise definition. It is possible to formally write down the mean-field equation corresponding to (1.4):

$$\begin{cases} d\bar{X}_t = -(\bar{X}_t - \mathcal{M}_\beta(\bar{\rho}_t)) dt + \sqrt{2\lambda^{-1}\mathcal{C}_\beta(\bar{\rho}_t)} dW_t, \\ \bar{\rho}_t = \text{Law}(\bar{X}_t). \end{cases} \quad (1.5)$$

When the parameters satisfy  $\lambda = (1 + \beta)^{-1}$ , the interacting particle system (1.4) enables to generate approximate samples from the probability distribution with density proportional to  $e^{-f}$ , that is from

$$\rho_f := \frac{e^{-f}}{\mathcal{Z}}, \quad \mathcal{Z} := \int_{\mathbf{R}^d} e^{-f(x)} dx.$$

More precisely, the authors of [12] show that for uniformly convex  $f$  and Gaussian initial data  $\bar{\rho}_0$ , the mean-field distribution  $\bar{\rho}_t$  converges exponentially in Wasserstein distance as  $t \rightarrow \infty$  to a Gaussian  $\mathcal{N}(m_\infty, C_\infty)$  that can be made arbitrarily close to the Laplace approximation of  $\rho_f$  by choosing a sufficiently large parameter  $\beta$ . In addition, it was noticed in [12] that, with  $\lambda := 1$ , the dynamics (1.4) behaves as a global optimization method for finding the global minimum of  $f: \mathbf{R}^d \rightarrow \mathbf{R}$ , just like the Consensus-Based Optimization method proposed in [46]. Since the convergence analysis in [12] is confined to the mean-field level, it is of great interest to prove rigorously that the particle dynamics (1.4) converges in an appropriate sense to the corresponding

mean-field dynamics (1.5) as  $J \rightarrow \infty$ . This has so far been left as an open problem. Recently, consensus-based sampling was extended to enable sampling of multi-modal distributions in [10], and for rare event estimation in [2].

### 1.3 Propagation of chaos

The CBO (1.1) and CBS (1.4) dynamics can both be recast as interacting particle systems of the form

$$dX_t^{(j)} = b\left(X_t^{(j)}, \mu_t^J\right) dt + \sigma\left(X_t^{(j)}, \mu_t^J\right) dW^{(j)}, \quad j \in \llbracket 1, J \rrbracket. \quad (1.6)$$

with drift coefficient  $b$  and diffusion coefficient  $\sigma$  depending on  $\mu$  through a convolution with appropriate kernels  $K_b$  and  $K_\sigma$ :

$$b(x, \mu) = \tilde{b}(x, K_b \star \mu), \quad \sigma(x, \mu) = \tilde{\sigma}(x, K_\sigma \star \mu). \quad (1.7)$$

Although the consensus-based dynamics can be written in the form (1.6), it is not possible to do so in such a way that  $\tilde{b}, \tilde{\sigma}, K_b$  and  $K_\sigma$  are all globally Lipschitz continuous and at the same time  $K_b$  and  $K_\sigma$  are globally bounded. Therefore, the classical McKean's theorem [16, Theorem 3.1] is not directly applicable, and a different approach is needed. We briefly discuss hereafter the existing mean-field results for CBO, noting that no results have been obtained so far for CBS.

- In [35], propagation of chaos for CBO is proved through a compactness argument based on Prokhorov's theorem. However, this method comes without convergence rates in terms of the number of particles  $J$ .
- Later, propagation of chaos by means of synchronous coupling was shown with optimal rates in [22] for a variant of CBO constrained to hypersurfaces.
- Recent results in [24] also show a probabilistic mean-field approximation of the form

$$\sup_{t \in [0, T]} \mathbf{E} \left[ \left| X_t^{(j)} - \bar{X}_t^{(j)} \right|^2 \mathbf{1}_{\Omega \setminus \Omega_M} \right] \leq \frac{C}{J}, \quad \Omega_M := \left\{ \sup_{t \in [0, T]} \frac{1}{J} \sum_{j=1}^J \max \left\{ \left| X_t^{(j)} \right|^4, \left| \bar{X}_t^{(j)} \right|^4 \right\} \geq M \right\}.$$

It is additionally proved in [24], by application of Markov's inequality, that the set of realizations  $\Omega_M$  has probability close to 0 for large  $M$ . Specifically, it holds that  $\mathbf{P}(\Omega_M) = \mathcal{O}(M^{-1})$ .

- Shortly before completing this work, we found out that propagation of chaos for CBO had been proved using a coupling approach very recently in [37]. In that work, the lack of global Lipschitz continuity of the weighted mean is handled by using stopping times, similarly to the approach we present in Subsection 5.1. There are, however, several differences between our results and those presented in [37]. Apart from the assumptions, which are slightly less restrictive in our work, the main technical differences are that (i) we improve existing Wasserstein stability estimates on the weighted mean (Corollary 3.3), (ii) we rely on a powerful result from the statistics literature [1, 19] to bound one of the terms in Sznitman's argument (Lemma 3.7) and (iii) we use a non-asymptotic central limit result to bound the probability that empirical measures exit a compact set (Subsection 2.3). Combined together, these ingredients enable to obtain mean-field result with the usual algebraic rate of convergence with respect to  $J$ , whereas the error scales as  $\ln(\ln(J))^{-1}$  in the mean-field result obtained in [37].

### 1.4 Our contributions

The main contributions of this paper are the following.

- We obtain quantitative results on the mean-field limit for CBO and CBS. More precisely, we prove results with an explicit rate of convergence with respect to the number  $J$  of particles. Investigating the roles of dimension and other problem parameters on the value of the constant prefactor is left for future work.
- We present two approaches for dealing with the weighted mean (in the case of CBO) and the weighted covariance (in the case of CBS). One of these approaches is based on the careful use of indicator sets, whereas the other is based on stopping times. The latter approach is inspired by a method initially proposed in [34] in order to handle local Lipschitz continuity issues in the context of numerical analysis for the Euler–Maruyama method.
- In order to prove mean-field results, we generalize existing theorems and prove new results on the well-posedness of the particle systems and their mean-field limit. We also present novel stability estimates for the weighted mean and the weighted covariance.

The rest of this paper is organized as follows. In [Section 2](#), we present our main results for the consensus-based dynamics (1.1) and (1.4). They are based on a number of auxiliary results, which are stated precisely and proved in [Section 3](#). These include Wasserstein stability estimates for the weighted mean and the weighted covariance, as well as moment estimates for the particle systems. In [Section 4](#), we prove well-posedness results for the particle systems and their mean-field limits. [Section 5](#) is reserved for extensions and perspectives for future work. Additional technical results are presented in [Appendix A](#). The interdependence of the sections is illustrated in [Figure 1](#).

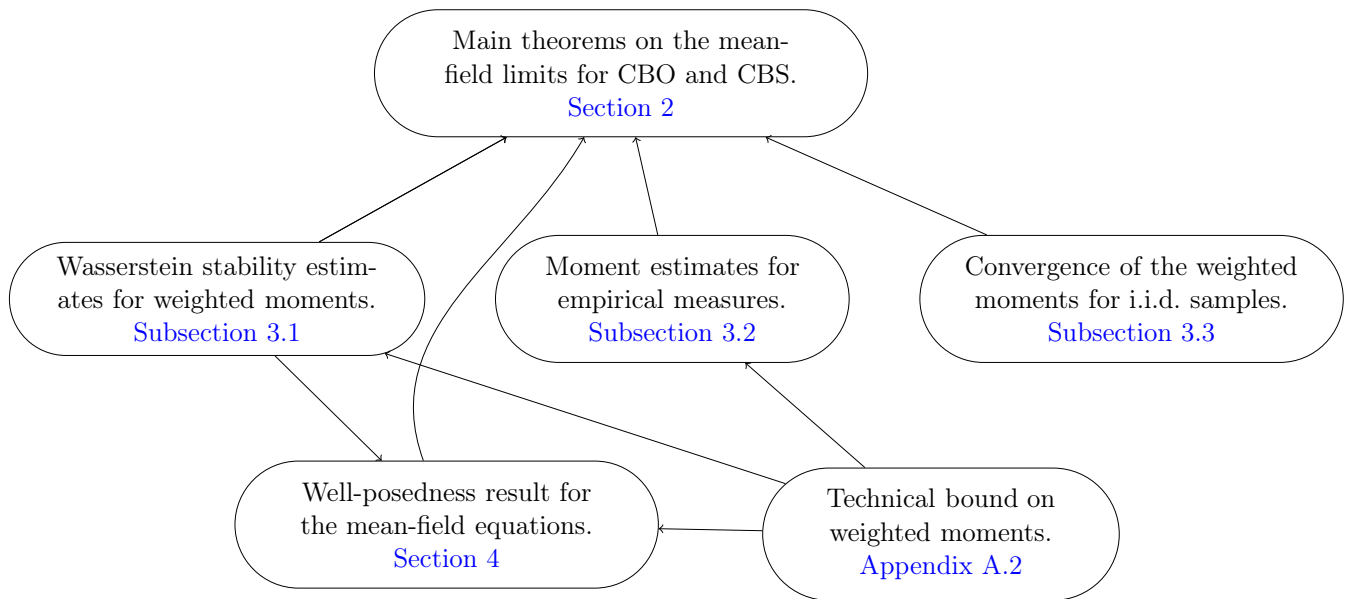


Figure 1: Interdependence of the results contained in this paper.

## 1.5 Notation

We let  $\mathbf{N} := \{0, 1, 2, 3, \dots\}$  and  $\mathbf{N}^+ := \{1, 2, 3, \dots\}$ . For symmetric matrices  $X$  and  $Y$ , the notation  $X \succcurlyeq Y$  means that  $X - Y$  is positive semidefinite, and the notation  $X \succ Y$  means that  $X - Y$  is strictly positive definite. The notation  $\mathbf{I}_d \in \mathbf{R}^{d \times d}$  refers to the identity matrix. The Euclidean norm in  $\mathbf{R}^d$  is denoted by  $|\cdot|$ . For a matrix  $X \in \mathbf{R}^{d \times d}$ , the notation  $\|X\|$  refers to the operator norm induced by the Euclidean vector norm, and the notation  $\|X\|_F$  refers to the Frobenius norm. The set of probability measures over a metric space  $E$  is denoted by  $\mathcal{P}(E)$ , and  $\mathcal{P}_p(E) \subset \mathcal{P}(E)$  denotes the set of probability measures with finite moments

up to order  $p$ , and  $\mathcal{P}_{p,R}(E) \subset \mathcal{P}_p(E)$  denotes the set of probability measures  $\mu$  such that

$$\left( \int_E |x|^p \mu(\mathrm{d}x) \right)^{\frac{1}{p}} \leq R.$$

Note that  $\mathcal{P}_{q,R}(E) \subset \mathcal{P}_{p,R}(E)$  for all  $0 \leq p < q$ . If the metric space  $E$  is not specified, then it is implicitly assumed that  $E = \mathbf{R}^d$ . For example, we sometimes write  $\mathcal{P}_{p,R}$  as a shorthand notation for  $\mathcal{P}_{p,R}(\mathbf{R}^d)$ . For a probability measure  $\mu \in \mathcal{P}(\mathbf{R}^d)$  we use the following notation for the mean and covariance:

$$\mathcal{M}(\mu) = \int_{\mathbf{R}^d} x \mu(\mathrm{d}x), \quad \mathcal{C}(\mu) = \int_{\mathbf{R}^d} (x - \mathcal{M}(\mu)) \otimes (x - \mathcal{M}(\mu)) \mu(\mathrm{d}x). \quad (1.8)$$

The notations  $\mathcal{M}_\beta(\mu)$  and  $\mathcal{C}_\beta(\mu)$  denote respectively the mean and covariance after reweighting by  $e^{-\beta f}$ . More precisely, we use the notation

$$\mathcal{M}_\beta(\mu) = \mathcal{M}(L_\beta \mu), \quad \mathcal{C}_\beta(\mu) = \mathcal{C}(L_\beta \mu), \quad L_\beta \mu := \frac{e^{-\beta f} \mu}{\int e^{-\beta f} \mathrm{d}\mu}. \quad (1.9)$$

For a probability measure  $\mu \in \mathcal{P}(E)$  and a function  $f: E \rightarrow \mathbf{R}$ , we use the short-hand notation

$$\mu[f] = \int_E f(x) \mu(\mathrm{d}x).$$

By a slight abuse of notation, we sometimes write  $\mu[f(x)]$  instead of  $\mu[f]$  for convenience. For example, for a probability measure  $\mu \in \mathcal{P}_2(\mathbf{R})$ , the notation  $\mu[x^2]$  refers to the second raw moment of  $\mu$ . Furthermore, for  $J \in \mathbf{N}^+$  we denote by  $\mu^J := \frac{1}{J} \sum_{j=1}^J \delta_{X^{(j)}}$  and  $\bar{\mu}^J := \frac{1}{J} \sum_{j=1}^J \delta_{\bar{X}^{(j)}}$  the empirical measures of the particles  $X^{(1)}, \dots, X^{(J)}$  and  $\bar{X}^{(1)}, \dots, \bar{X}^{(J)}$ , respectively. Finally, throughout this paper, the notation  $\Omega$  refers to the sample space, and the notation  $C$  refers to a constant whose exact value is irrelevant in the context and may change from line to line.

## 2 Main results

This section is organized as follows. In [Subsection 2.1](#), we present the assumptions used throughout this work. Then, after stating well-posedness results for CBO and CBS in [Subsection 2.2](#), we prove in [Subsection 2.3](#) a key preparatory result for the mean-field limits. We then prove quantitative mean-field results for Consensus-Based Optimization and Sampling in [Subsections 2.4](#) and [2.5](#), respectively.

### 2.1 Main assumptions

We assume throughout this paper that the diffusion operator  $S: \mathbf{R}^d \rightarrow \mathbf{R}^{d \times d}$  appearing in the CBO dynamics [\(1.1\)](#) is globally Lipschitz continuous. For the objective function  $f$ , we use the following assumption.

**Assumption H.** *We assume the following properties for the target function  $f$ .*

**(H1)** *The function  $f: \mathbf{R}^d \rightarrow \mathbf{R}$  is bounded from below by  $f_\star = \inf f$  and satisfies*

$$\forall x, y \in \mathbf{R}^d, \quad |f(x) - f(y)| \leq L_f (1 + |x| + |y|)^s |x - y|, \quad (2.1)$$

*for some constants  $L_f > 0$  and  $s \geq 0$ .*

(H2) There exist constants  $c_\ell, c_u > 0$  and  $u \geq \ell \geq 0$  such that

$$\forall x \in \mathbf{R}^d, \quad f(x) - f_\star \leq c_u(|x|^u + 1), \quad (2.2a)$$

$$\forall x \in \mathbf{R}^d, \quad f(x) - f_\star \geq c_\ell(|x|^\ell - 1). \quad (2.2b)$$

**Remark 2.1.** A few remarks are in order.

- When  $\ell = u = 0$ , assumption (H2) implies that  $f$  is bounded from above.
- Observe that (H1) implies (2.2a) with  $u = s + 1$ .
- In the particular case where  $s = 1$  and either  $u = \ell = 0$  or  $u = \ell = 2$ , the assumptions in Assumption H coincide with those in [11]. The well-posedness and mean-field results proved in this paper generalize beyond this setting, as explained below.

In the rest of this paper, we denote by  $\mathcal{A}(s, \ell, u)$  the set of objective functions that satisfy Assumption H with these parameters. Our main results hold only for  $f \in \mathcal{A}(s, \ell, \ell)$ , i.e. under Assumption H with  $\ell = u$ . Many of the results we present hold only for values of  $p$  in a range dictated by the interval of validity of the Wasserstein stability estimates presented in Subsection 3.1. In order to avoid repetitions, we introduce the associated notation

$$p_{\mathcal{M}}(s, \ell) := \begin{cases} s + 2, & \text{if } \ell = 0, \\ 1, & \text{if } \ell > 0, \end{cases} \quad p_{\mathcal{C}}(s, \ell) := \begin{cases} s + 3, & \text{if } \ell = 0, \\ 1, & \text{if } \ell > 0. \end{cases} \quad (2.3)$$

## 2.2 Well-posedness for CBO and CBS

In this section, we state well-posedness results for CBO and CBS, referring for the proofs to Section 4. While for CBO, the following theorems are only generalizations of already existing results [11], the well-posedness proofs for the CBS particle and mean-field systems are new.

**Theorem 2.2** (Existence and uniqueness for the particle models). *Assume that  $f: \mathbf{R}^d \rightarrow \mathbf{R}$  is locally Lipschitz continuous and fix  $J \in \mathbf{N}^+$ . Then, the stochastic differential equations (1.1) and (1.4) have unique strong solutions  $\{\mathbf{X}_t^{(J)} \mid t \geq 0\}$  for any initial condition  $\mathbf{X}_0^{(J)}$  which is independent of the Brownian motions  $\mathbf{W}_t = (W_t^{(1)}, \dots, W_t^{(J)})$ . The solutions are almost surely continuous.*

In the case of CBO, Theorem 2.2 was already proved in [11, Theorem 2.1]. We now present well-posedness results for the mean-field SDEs, beginning with CBO. For  $f \in \mathcal{A}(1, 0, 0) \cup \mathcal{A}(1, 2, 2)$ , the following result generalizes [11, Theorems 3.1 and 3.2].

**Theorem 2.3** (CBO: Existence and uniqueness for the mean-field SDE). *Suppose that  $f \in \mathcal{A}(s, \ell, \ell)$  with parameters  $s \geq 0, \ell \geq 0$ . Assume also that  $\bar{\rho}_0 \in \mathcal{P}_p(\mathbf{R}^d)$  for  $p \geq 2 \vee p_{\mathcal{M}}(s, \ell)$ . Then, for each final time  $T > 0$ , there exists a unique process  $\bar{X} \in C([0, T], \mathbf{R}^d)$  satisfying (1.3) in the strong sense with initial condition  $\bar{X}_0 = x_0$ . Furthermore, it holds that*

$$\mathbf{E} \left[ \sup_{t \in [0, T]} |\bar{X}_t|^p \right] < \infty \quad \text{and} \quad \sup_{t \in [0, T]} |\mathcal{M}_\beta(\bar{\rho}_t)| < \infty, \quad \bar{\rho}_t := \text{Law}(\bar{X}_t), \quad (2.4)$$

and the function  $t \mapsto \mathcal{M}_\beta(\bar{\rho}_t)$  is continuous over  $[0, T]$ . Finally, the function  $t \mapsto \bar{\rho}_t$  belongs to  $C([0, T], \mathcal{P}_p(\mathbf{R}^d))$  and satisfies the following non-local Fokker–Planck equation in the weak sense:

$$\partial_t \bar{\rho} = \nabla \cdot \left( (x - \mathcal{M}_\beta(\bar{\rho})) \bar{\rho} \right) + \frac{1}{2} \nabla \cdot \nabla \cdot \left( D(\bar{\rho}, x) \bar{\rho} \right). \quad (2.5)$$



where  $D(\bar{\rho}, x) := S(x - \mathcal{M}_\beta(\bar{\rho}))S(x - \mathcal{M}_\beta(\bar{\rho}))^\top$ .

**Theorem 2.4** (CBS: Existence and uniqueness for the mean-field SDE). *Suppose that  $f \in \mathcal{A}(s, \ell, \ell)$  with parameters  $s \geq 0, \ell \geq 0$ . Assume also that  $\bar{\rho}_0 \in \mathcal{P}_p(\mathbf{R}^d)$  for  $p \geq 2\vee pc(s, \ell)$ , that  $C_\beta(\bar{\rho}_0) \succ 0$ , and fix  $x_0 \sim \bar{\rho}_0$ . Then, for each final time  $T > 0$ , there exists a unique process  $\bar{X} \in C([0, T], \mathbf{R}^d)$  satisfying (1.5) in the strong sense with initial condition  $\bar{X}_0 = x_0$ . Furthermore, it holds that*

$$\mathbf{E} \left[ \sup_{t \in [0, T]} |\bar{X}_t|^p \right] < \infty, \quad \sup_{t \in [0, T]} |\mathcal{M}_\beta(\bar{\rho}_t)| < \infty, \quad \sup_{t \in [0, T]} \|C_\beta(\bar{\rho}_t)\| < \infty, \quad \bar{\rho}_t := \text{Law}(\bar{X}_t), \quad (2.6)$$

and the functions  $t \mapsto \mathcal{M}_\beta(\bar{\rho}_t)$ ,  $t \mapsto C_\beta(\bar{\rho}_t)$  are continuous over  $[0, T]$ . Finally, the function  $t \mapsto \bar{\rho}_t$  belongs to  $C([0, T], \mathcal{P}_p(\mathbf{R}^d))$  and satisfies the following non-local Fokker–Planck equation in the weak sense:

$$\partial_t \bar{\rho} = \nabla \cdot \left( (x - \mathcal{M}_\beta(\bar{\rho}))\bar{\rho} + \lambda^{-1} C_\beta(\bar{\rho}) \nabla \bar{\rho} \right). \quad (2.7)$$

### 2.3 Key preparatory lemma

We first show a key preparatory lemma which is needed for the mean-field results. This result is a simple corollary of Markov’s inequality,

**Lemma 2.5** (Bound on the probability of large excursions). *Let  $(Z_j)_{j \in \mathbf{N}^+}$  be a family of i.i.d.  $\mathbf{R}$ -valued random variables such that  $\mathbf{E}[|Z_1|^r] < \infty$  for some  $r \geq 2$ . Then for all  $R > \mathbf{E}[|Z_1|]$ , there exists a constant  $C > 0$  such that*

$$\forall J \in \mathbf{N}^+, \quad \mathbf{P} \left[ \frac{1}{J} \sum_{j=1}^J Z_j \geq R \right] \leq C J^{-\frac{r}{2}}.$$

*Proof.* This follows from a generalization of Chebychev’s inequality [44, Exercise 3.21]. Let

$$X = \frac{1}{J} \sum_{j=1}^J Z_j.$$

By the classical Marcinkiewicz–Zygmund inequality, it holds that

$$\begin{aligned} \mathbf{E}|X - \mathbf{E}[X]|^r &\leq J^{-r} C_{\text{MZ}}(r) \mathbf{E} \left[ \left( \sum_{j=1}^J |Z_j - \mathbf{E}[Z_j]|^2 \right)^{\frac{r}{2}} \right] \\ &\leq J^{-\frac{r}{2}} C_{\text{MZ}}(r) \mathbf{E} \left[ \frac{1}{J} \sum_{j=1}^J |Z_j - \mathbf{E}[Z_j]|^r \right] = J^{-\frac{r}{2}} C_{\text{MZ}}(r) \mathbf{E} \left[ |Z_1 - \mathbf{E}[Z_1]|^r \right], \end{aligned}$$

where we used the convexity of the function  $x \mapsto x^{\frac{r}{2}}$  in the second inequality. Therefore, using the Markov inequality, we deduce that

$$\mathbf{P}[X \geq R] \leq \mathbf{P}[|X - \mathbf{E}[X]|^r \geq (R - \mathbf{E}[X])^r] \leq \mathbf{E} \left[ \frac{|X - \mathbf{E}[X]|^r}{(R - \mathbf{E}[X])^r} \right] \leq \frac{C J^{-\frac{r}{2}}}{(R - \mathbf{E}[X])^r},$$

which concludes the proof.  $\square$

### 2.4 Quantitative mean-field limit for CBO

We now present and prove the mean-field results, starting with CBO. Synchronous coupling for McKean–Vlasov diffusions with drift  $b: \mathbf{R}^d \times \mathcal{P}(\mathbf{R}^d) \rightarrow \mathbf{R}^d$  and diffusion  $\sigma: \mathbf{R}^d \times \mathcal{P}(\mathbf{R}^d) \rightarrow \mathbf{R}^{d \times d}$  assumes the following

setting [15, 16]. Given i.i.d. Brownian motions  $(W^{(j)})_{j \in \mathbf{N}^+}$  and initial conditions  $(X_0^{(j)})_{j \in \mathbf{N}^+}$  sampled i.i.d. from some  $\bar{\rho}_0 \in \mathcal{P}(\mathbf{R}^d)$ , we consider for each  $J \in \mathbf{N}^+$  the interacting particle system

$$\forall j \in \llbracket 1, J \rrbracket, \quad X_t^{(j)} = X_0^{(j)} + \int_0^t b(X_s^{(j)}, \mu_s^J) ds + \int_0^t \sigma(X_s^{(j)}, \mu_s^J) dW_s^{(j)}, \quad (2.8)$$

to which we couple the system of i.i.d. mean-field McKean-Vlasov diffusions

$$\forall N \in \mathbf{N}^+, \forall j \in \llbracket 1, J \rrbracket, \quad \begin{cases} \bar{X}_t^{(j)} = X_0^{(j)} + \int_0^t b(\bar{X}_s^{(j)}, \bar{\rho}_s) ds + \int_0^t \sigma(\bar{X}_s^{(j)}, \bar{\rho}_s) dW_s^{(j)}, \\ \bar{\rho}_t = \text{Law}(\bar{X}_t^{(j)}), \end{cases} \quad (2.9)$$

driven by the same Brownian motions and initial conditions as (2.8). In the case of CBO, the drift and diffusion coefficients for (2.8) and (2.9) are given by

$$b(x, \mu) := -(x - \mathcal{M}_\beta(\mu)) \quad \text{and} \quad \sigma(x, \mu) := S(x - \mathcal{M}_\beta(\mu)), \quad (2.10)$$

where the diffusion operator  $S: \mathbf{R}^d \rightarrow \mathbf{R}^{d \times d}$  is, as always in this work, a globally Lipschitz continuous function.

**Theorem 2.6** (Mean-field limit for CBO). *Suppose that  $f \in \mathcal{A}(s, \ell, \ell)$  with  $s, \ell \geq 0$  and  $\bar{\rho}_0 \in \mathcal{P}(\mathbf{R}^d)$  has moments of sufficiently high order. Consider the systems (2.8) and (2.9) with the coefficients given by (2.10). Then, for each  $p > 0$ , there exists a constant  $C > 0$  independent of  $J$  such that*

$$\forall J \in \mathbf{N}^+, \quad \forall j \in \llbracket 1, J \rrbracket, \quad \left( \mathbf{E} \left[ \sup_{t \in [0, T]} |X_t^{(j)} - \bar{X}_t^{(j)}|^p \right] \right)^{\frac{1}{p}} \leq C J^{-\frac{1}{2}}. \quad (2.11)$$

**Remark 2.7.** *We prove below that the following sharper statement holds. Suppose that  $f \in \mathcal{A}(s, \ell, \ell)$  with  $s, \ell \geq 0$  and that  $\bar{\rho}_0 \in \mathcal{P}_q(\mathbf{R}^d)$ , and consider the systems (2.8) and (2.9) with the coefficients given by (2.10). If  $q \geq \max\{4, 2p_{\mathcal{M}}(s, \ell)\}$ , where  $p_{\mathcal{M}}$  is given in (2.3), then for all  $p \in (0, \frac{q}{2}]$  there is  $C > 0$  independent of  $J$  such that*

$$\forall J \in \mathbf{N}^+, \quad \forall j \in \llbracket 1, J \rrbracket, \quad \left( \mathbf{E} \left[ \sup_{t \in [0, T]} |X_t^{(j)} - \bar{X}_t^{(j)}|^p \right] \right)^{\frac{1}{p}} \leq C J^{-\min\left\{\frac{1}{2}, \frac{q-p}{2p^2}, \frac{q-(2\sqrt{p}\mathcal{M})}{2(2\sqrt{p}\mathcal{M})^2}\right\}}. \quad (2.12)$$

**Remark 2.8.** *Theorem 2.6 implies pathwise Wasserstein- $p$  chaos. More precisely, it holds that*

$$W_p \left( \rho_{[0, T]}^J, \bar{\rho}_{[0, T]}^{\otimes J} \right) \leq C J^{-\frac{1}{2}}.$$

where  $\rho_{[0, T]}^J$  is the pathwise law of the microscopic particle system over  $[0, T]$  and  $\bar{\rho}_{[0, T]}$  is the pathwise mean-field law over  $[0, T]$ . That is to say,

$$\begin{cases} \rho_{[0, T]}^J := \text{Law} \left( X^{(1)}, \dots, X^{(J)} \right) \in \mathcal{P} \left( C([0, T], \mathbf{R}^d)^J \right), \\ \bar{\rho}_{[0, T]} := \text{Law}(\bar{X}) \in \mathcal{P} \left( C([0, T], \mathbf{R}^d) \right). \end{cases}$$

Here,  $W_p$  denotes the pathwise Wasserstein- $p$  distance, associated with the following metric on  $\left( C([0, T], \mathbf{R}^d) \right)^J$ :

$$d(\mathbf{X}, \mathbf{Y}) = \sup_{t \in [0, T]} \frac{1}{J} \sum_{j=1}^J |X_t^{(j)} - Y_t^{(j)}|, \quad \mathbf{X}, \mathbf{Y} \in \left( C([0, T], \mathbf{R}^d) \right)^J.$$

See [15, Definitions 3.1 and 3.5] for more details.

Before presenting the proof, we summarize below a number of useful auxiliary results which are valid for  $f \in \mathcal{A}(s, \ell, \ell)$ . These results are stated precisely and proved in Section 3.

- (a) **Wasserstein stability estimate for the weighted mean.** We prove in Corollary 3.3 that, for all  $R > 0$  and for all  $p \geq p_{\mathcal{M}}(s, \ell)$ , where  $p_{\mathcal{M}}(s, \ell) \in \mathbf{R}^+$  is given in (2.3), there exists  $\tilde{L}_{\mathcal{M}} = \tilde{L}_{\mathcal{M}}(R, p) > 0$  such that

$$\forall (\mu, \nu) \in \mathcal{P}_{p,R}(\mathbf{R}^d) \times \mathcal{P}_p(\mathbf{R}^d), \quad |\mathcal{M}_{\beta}(\mu) - \mathcal{M}_{\beta}(\nu)| \leq \tilde{L}_{\mathcal{M}} W_p(\mu, \nu).$$

- (b) **Moment estimate for the particle system.** We prove in Lemma 3.5 that, if the initial condition satisfies  $\bar{\rho}_0 \in \mathcal{P}_q(\mathbf{R}^d)$ , then there exists a constant  $\kappa > 0$  independent of  $J$  such that

$$\forall J \in \mathbf{N}^+, \quad \mathbf{E} \left[ \sup_{t \in [0, T]} |X_t^{(j)}|^q \right] \vee \mathbf{E} \left[ \sup_{t \in [0, T]} |\mathcal{M}_{\beta}(\mu_t^J)|^q \right] \leq \kappa.$$

- (c) **Convergence of the weighted mean for i.i.d. samples.** We prove in Lemma 3.7 that, for any reals  $r > p > 0$  and any  $\mu \in \mathcal{P}_r(\mathbf{R}^d)$ , there is  $C$  such that

$$\mathbf{E} |\mathcal{M}_{\beta}(\bar{\mu}^J) - \mathcal{M}_{\beta}(\mu)|^p \leq C J^{-\frac{p}{2}}, \quad \bar{\mu}^J := \frac{1}{J} \sum_{j=1}^J \delta_{\bar{X}^{(j)}}, \quad \{\bar{X}^{(j)}\}_{j \in \mathbf{N}} \stackrel{\text{i.i.d.}}{\sim} \mu.$$

*Proof of Theorem 2.6.* As mentioned in Remark 2.7, we prove the sharp result (2.12). It is sufficient to prove this statement for all  $p \in [2 \vee p_{\mathcal{M}}, \frac{q}{2}]$  (by the assumption on  $q$ , this interval is nonempty). Indeed, if the statement holds true in this case, then by Jensen's inequality, it holds for all  $p \in (0, 2 \vee p_{\mathcal{M}}]$  that

$$\begin{aligned} \left( \mathbf{E} \left[ \sup_{t \in [0, T]} |X_t^{(j)} - \bar{X}_t^{(j)}|^p \right] \right)^{\frac{1}{p}} &\leq \left( \mathbf{E} \left[ \sup_{t \in [0, T]} |X_t^{(j)} - \bar{X}_t^{(j)}|^{2 \vee p_{\mathcal{M}}} \right] \right)^{\frac{1}{2 \vee p_{\mathcal{M}}}} \\ &\leq C J^{-\min\left\{\frac{1}{2}, \frac{q-2 \vee p_{\mathcal{M}}}{2(2 \vee p_{\mathcal{M}})^2}\right\}} = C J^{-\min\left\{\frac{1}{2}, \frac{q-p}{2p^2}, \frac{q-2 \vee p_{\mathcal{M}}}{2(2 \vee p_{\mathcal{M}})^2}\right\}}, \end{aligned}$$

where we used that the function  $z \mapsto (q-z)/z^2$  is decreasing. Take  $p \in [2 \vee p_{\mathcal{M}}, q/2]$  and fix  $j \in \llbracket 1, J \rrbracket$ . As in Sznitman's argument [16, Theorem 3.1], we begin by writing

$$\begin{aligned} |X_t^{(j)} - \bar{X}_t^{(j)}|^p &\leq 2^{p-1} \left| \int_0^t b(X_s^{(j)}, \mu_s^J) - b(\bar{X}_s^{(j)}, \bar{\rho}_s) \, ds \right|^p \\ &\quad + 2^{p-1} \left| \int_0^t \sigma(X_s^{(j)}, \mu_s^J) - \sigma(\bar{X}_s^{(j)}, \bar{\rho}_s) \, dW_s \right|^p. \end{aligned}$$

By the Burkholder–Davis–Gundy inequality, see [43, Theorem 7.3], it holds for any  $t \in [0, T]$  that

$$\begin{aligned} \mathbf{E} \left[ \sup_{s \in [0, t]} |X_s^{(j)} - \bar{X}_s^{(j)}|^p \right] &\leq 2^{p-1} \mathbf{E} \left( \int_0^t |b(X_s^{(j)}, \mu_s^J) - b(\bar{X}_s^{(j)}, \bar{\rho}_s)| \, ds \right)^p \\ &\quad + C_{\text{BDG}} 2^{p-1} \mathbf{E} \left( \int_0^t \left\| \sigma(X_s^{(j)}, \mu_s^J) - \sigma(\bar{X}_s^{(j)}, \bar{\rho}_s) \right\|_{\mathbb{F}}^2 \, ds \right)^{\frac{p}{2}} \\ &\leq (2T)^{p-1} \mathbf{E} \int_0^t |b(X_s^{(j)}, \mu_s^J) - b(\bar{X}_s^{(j)}, \bar{\rho}_s)|^p \, ds \\ &\quad + C_{\text{BDG}} 2^{p-1} T^{\frac{p}{2}-1} \mathbf{E} \int_0^t \left\| \sigma(X_s^{(j)}, \mu_s^J) - \sigma(\bar{X}_s^{(j)}, \bar{\rho}_s) \right\|_{\mathbb{F}}^p \, ds. \end{aligned} \quad (2.13)$$

By the triangle inequality, the first expectation on the right-hand side can be bounded as follows:

$$\begin{aligned} \mathbf{E} \int_0^t \left| b \left( X_s^{(j)}, \mu_s^J \right) - b \left( \bar{X}_s^{(j)}, \bar{\rho}_s \right) \right|^p ds &\leq 2^{p-1} \mathbf{E} \int_0^t \left| b \left( X_s^{(j)}, \mu_s^J \right) - b \left( \bar{X}_s^{(j)}, \bar{\mu}_s^J \right) \right|^p ds \\ &\quad + 2^{p-1} \mathbf{E} \int_0^t \left| b \left( \bar{X}_s^{(j)}, \bar{\mu}_s^J \right) - b \left( \bar{X}_s^{(j)}, \bar{\rho}_s \right) \right|^p ds. \end{aligned} \quad (2.14)$$

**Bounding the first term in (2.14).** From (2.10) we have

$$\begin{aligned} \mathbf{E} \left| b \left( X_s^{(j)}, \mu_s^J \right) - b \left( \bar{X}_s^{(j)}, \bar{\mu}_s^J \right) \right|^p &= \mathbf{E} \left| \left( \bar{X}_s^{(j)} - X_s^{(j)} + \mathcal{M}_\beta \left( \mu_s^J \right) - \mathcal{M}_\beta \left( \bar{\mu}_s^J \right) \right) \right|^p \\ &\leq 2^{p-1} \mathbf{E} \left| X_s^{(j)} - \bar{X}_s^{(j)} \right|^p + 2^{p-1} \mathbf{E} \left| \mathcal{M}_\beta \left( \mu_s^J \right) - \mathcal{M}_\beta \left( \bar{\mu}_s^J \right) \right|^p. \end{aligned} \quad (2.15)$$

In order to deal with the fact that the weighted mean does not have global Lipschitz properties in the sense of McKean's framework [16, Theorem 3.1], we introduce for each  $t \in [0, T]$  and  $J \in \mathbf{N}^+$  the set

$$\Omega_{J,t} := \left\{ \omega \in \Omega : \frac{1}{J} \sum_{j=1}^J \left| \bar{X}_t^{(j)}(\omega) \right|^p \geq R \right\}, \quad (2.16)$$

where  $R > \mathbf{E}[Z_j]$  is fixed. Define

$$Z_j = \sup_{t \in [0, T]} \left| \bar{X}_t^{(j)} \right|^p.$$

By Theorem 2.3 it holds that  $\mathbf{E} [|Z_j|^{q/p}] < \infty$ . Note that  $q \geq 2p$ . Therefore, by Lemma 2.5 there exists a constant  $C > 0$  such that

$$\forall t \in [0, T], \quad \forall J \in \mathbf{N}^+, \quad \mathbf{P} [\Omega_{J,t}] \leq \mathbf{P} \left[ \frac{1}{J} \sum_{j=1}^J Z_j \geq R \right] \leq C J^{-\frac{q}{2p}}.$$

We split the second term on the right-hand of (2.15) as follows:

$$\begin{aligned} \mathbf{E} \left| \mathcal{M}_\beta \left( \mu_s^J \right) - \mathcal{M}_\beta \left( \bar{\mu}_s^J \right) \right|^p &= \mathbf{E} \left[ \left| \mathcal{M}_\beta \left( \mu_s^J \right) - \mathcal{M}_\beta \left( \bar{\mu}_s^J \right) \right|^p \mathbf{1}_{\Omega \setminus \Omega_{J,s}} \right] \\ &\quad + \mathbf{E} \left[ \left| \mathcal{M}_\beta \left( \mu_s^J \right) - \mathcal{M}_\beta \left( \bar{\mu}_s^J \right) \right|^p \mathbf{1}_{\Omega_{J,s}} \right]. \end{aligned} \quad (2.17)$$

Let us now bound each term on the right-hand side of (2.17) separately.

- Since the  $p$ -th moment of  $\bar{\mu}_s^J$  is bounded from above by  $R$  for all  $\omega \in \Omega \setminus \Omega_{J,s}$ , we can use the stability estimate on the weighted mean given (see Item (a) and Corollary 3.3) to obtain

$$\mathbf{E} \left[ \left| \mathcal{M}_\beta \left( \mu_s^J \right) - \mathcal{M}_\beta \left( \bar{\mu}_s^J \right) \right|^p \mathbf{1}_{\Omega \setminus \Omega_{J,s}} \right] \leq C \mathbf{E} \left[ W_p \left( \mu_s^J, \bar{\mu}_s^J \right)^p \right]$$

for the first term in (2.17). Since in addition

$$\mathbf{E} \left[ W_p \left( \mu_s^J, \bar{\mu}_s^J \right)^p \right] \leq \mathbf{E} \left[ \frac{1}{J} \sum_{j=1}^J \left| X_s^{(j)} - \bar{X}_s^{(j)} \right|^p \right] = \mathbf{E} \left| X_s^{(j)} - \bar{X}_s^{(j)} \right|^p,$$

it follows that

$$\mathbf{E} \left[ \left| \mathcal{M}_\beta \left( \mu_s^J \right) - \mathcal{M}_\beta \left( \bar{\mu}_s^J \right) \right|^p \mathbf{1}_{\Omega \setminus \Omega_{J,s}} \right] \leq C \mathbf{E} \left| X_s^{(j)} - \bar{X}_s^{(j)} \right|^p.$$

- We now bound the second term in (2.17). From Item (b) (see Lemma 3.5) and (2.4) we have

$$\mathbf{E} \left[ |\mathcal{M}_\beta(\mu_s^J) - \mathcal{M}_\beta(\bar{\mu}_s^J)|^q \right] \leq 2^{q-1} \left( \mathbf{E} |\mathcal{M}_\beta(\mu_s^J)|^q + \mathbf{E} |\mathcal{M}_\beta(\bar{\mu}_s^J)|^q \right) \leq 2^q \kappa.$$

Therefore, we obtain by Hölder's inequality and Lemma 2.5 that

$$\mathbf{E} \left[ |\mathcal{M}_\beta(\mu_s^J) - \mathcal{M}_\beta(\bar{\mu}_s^J)|^p \mathbf{1}_{\Omega_{J,s}} \right] \leq \left( \mathbf{E} \left[ |\mathcal{M}_\beta(\mu_s^J) - \mathcal{M}_\beta(\bar{\mu}_s^J)|^q \right] \right)^{\frac{p}{q}} \left( \mathbf{P}[\Omega_{J,s}] \right)^{\frac{q-p}{q}} \leq 2^p \kappa^{\frac{p}{q}} J^{-\frac{q-p}{2p}}.$$

We have thus shown that

$$\mathbf{E} |\mathcal{M}_\beta(\mu_s^J) - \mathcal{M}_\beta(\bar{\mu}_s^J)|^p \leq C \mathbf{E} \left| X_s^{(j)} - \bar{X}_s^{(j)} \right|^p + C J^{-\frac{q-p}{2p}},$$

which in turn by (2.15) implies that

$$\mathbf{E} \left| b \left( X_s^{(j)}, \mu_s^J \right) - b \left( \bar{X}_s^{(j)}, \bar{\mu}_s^J \right) \right|^p \leq C \mathbf{E} \left| X_s^{(j)} - \bar{X}_s^{(j)} \right|^p + C J^{-\frac{q-p}{2p}}.$$

**Bounding the second term in (2.14).** For the second term on the right-hand side of (2.14), we have by Item (c) (see Lemma 3.7) that

$$\mathbf{E} \left| b \left( \bar{X}_s^{(j)}, \bar{\mu}_s^J \right) - b \left( \bar{X}_s^{(j)}, \bar{\rho}_s \right) \right|^p = \mathbf{E} |\mathcal{M}_\beta(\bar{\mu}_s^J) - \mathcal{M}_\beta(\bar{\rho}_s)|^p \leq C J^{-\frac{p}{2}}.$$

**Conclusion.** Combining the bounds on both terms in (2.14), we deduce

$$\mathbf{E} \left[ \int_0^t \left| b \left( X_s^{(j)}, \mu_s^J \right) - b \left( \bar{X}_s^{(j)}, \bar{\rho}_s \right) \right|^p ds \right] \leq C J^{-\min\left\{\frac{p}{2}, \frac{q-p}{2p}\right\}} + C \int_0^t \mathbf{E} \left[ \sup_{u \in [0,s]} \left| X_u^{(j)} - \bar{X}_u^{(j)} \right|^p \right] ds.$$

Using that  $S: \mathbf{R} \rightarrow \mathbf{R}^{d \times d}$  is globally Lipschitz continuous, we obtain analogously

$$\mathbf{E} \left[ \int_0^t \left\| \sigma \left( X_s^{(j)}, \mu_s^J \right) - \sigma \left( \bar{X}_s^{(j)}, \bar{\rho}_s \right) \right\|^p ds \right] \leq C J^{-\min\left\{\frac{p}{2}, \frac{q-p}{2p}\right\}} + C \int_0^t \mathbf{E} \left[ \sup_{u \in [0,s]} \left| X_u^{(j)} - \bar{X}_u^{(j)} \right|^p \right] ds.$$

Substituting in (2.13), we finally obtain

$$\mathbf{E} \left[ \sup_{s \in [0,t]} \left| X_s^{(j)} - \bar{X}_s^{(j)} \right|^p \right] \leq C J^{-\min\left\{\frac{p}{2}, \frac{q-p}{2p}\right\}} + C \int_0^t \mathbf{E} \left[ \sup_{u \in [0,s]} \left| X_u^{(j)} - \bar{X}_u^{(j)} \right|^p \right] ds.$$

The conclusion then follows from Grönwall's lemma, since  $\frac{q-p}{2p^2} \leq \frac{q-(2\sqrt{p\mathcal{M}})}{2(2\sqrt{p\mathcal{M}})^2}$  for the value of  $p$  considered.  $\square$

**Remark 2.9.** *The proof is similar to that of McKean's theorem [16, Theorem 3.1], with an important difference: we eliminate a set of small probability in order to apply the local stability estimate in Item (a) on its complement.*

**Remark 2.10.** *A crucial detail in the proof is that Item (a) only assumes that  $(\mu, \nu) \in \mathcal{P}_{p,R} \times \mathcal{P}_p$  and not  $(\mu, \nu) \in \mathcal{P}_{p,R} \times \mathcal{P}_{p,R}$ ; see Corollary 3.3 below. This allows us to define  $\Omega_{J,t}$  in (2.16) only in terms of the i.i.d. mean-field particles  $\bar{X}_t^{(j)}$  for which Lemma 2.5 can be applied; the event  $\frac{1}{J} \sum_{j=1}^J \left| X_t^{(j)} \right|^p \geq R$  need not be included in the definition of  $\Omega_{J,t}$ .*

**Remark 2.11.** *In his famous lecture notes [47], Sznitman shows for a laboratory example that*

$$\forall j \in \llbracket 1, J \rrbracket, \quad \mathbf{E} \left[ \sup_{t \in [0,T]} \left| X_t^{(j)} - \bar{X}_t^{(j)} \right|^p \right] \leq C J^{-\frac{p}{2}}.$$

More precisely, this statement is proved for  $p = 1$ , and the author notes that the result generalizes to higher moment, which can be achieved by using, among other technical adaptations of his argument, the Marcinkiewicz–Zygmund inequality [17]. Although CBO does not fit within Sznitman’s simplified framework, Item (c) enables to show a bound with the same rate for the second term on the right-hand side of (2.14),

$$\mathbf{E} \left[ \left| b \left( \overline{X}_s^{(j)}, \overline{\mu}_s^J \right) - b \left( \overline{X}_s^{(j)}, \overline{\rho}_s \right) \right|^p \mathbf{1}_{\Omega \setminus \Omega_{J,t}} \right] \leq \frac{C}{J^{\frac{p}{2}}}.$$

The limiting factor, for the overall rate of convergence in Theorem 2.6, is the bound on the probability  $\mathbf{P}[\Omega_{J,t}]$ .

## 2.5 Quantitative mean-field limit for CBS

Using the same reasoning as that in the proof of Theorem 2.6, we can prove a mean-field result for the CBS dynamics (1.4). We state the theorem and, given the similarity of the proof with that of Theorem 2.6, only list the key ingredients needed for the proof.

**Theorem 2.12** (Mean-field limit for CBS). *Suppose that  $f \in \mathcal{A}(s, \ell, \ell)$  with  $s, \ell \geq 0$  and that  $\overline{\rho}_0 \in \mathcal{P}(\mathbf{R}^d)$  with  $\mathcal{C}(\overline{\rho}_0) \succ 0$  has moments of sufficiently high order. Consider the systems (2.8) and (2.9) with the coefficients given by*

$$b(x, \mu) := -(x - \mathcal{M}_\beta(\mu)) \quad \text{and} \quad \sigma(x, \mu) := \sqrt{2\lambda^{-1}\mathcal{C}_\beta(\mu)}. \quad (2.18)$$

Then for all  $p > 0$ , there is  $C > 0$  independent of  $J$  such that

$$\forall J \in \mathbf{N}^+, \quad \forall j \in \llbracket 1, J \rrbracket, \quad \left( \mathbf{E} \left[ \sup_{t \in [0, T]} \left| X_t^{(j)} - \overline{X}_t^{(j)} \right|^p \right] \right)^{\frac{1}{p}} \leq C J^{-\frac{1}{2}}. \quad (2.19)$$

**Remark 2.13.** *In the same way as for CBO, it is possible to prove the following sharper statement. Suppose that  $f \in \mathcal{A}(s, \ell, \ell)$  with  $s, \ell \geq 0$  and that  $\overline{\rho}_0 \in \mathcal{P}_q(\mathbf{R}^d)$  with  $\mathcal{C}(\overline{\rho}_0) \succ 0$ . Consider the system (2.8) and (2.9) with the coefficients given by (2.18). If  $q \geq 4p_{\mathcal{M}}(s, \ell)$ , then for all  $p \in (0, \frac{q}{2}]$ , there is  $C > 0$  independent of  $J$  such that*

$$\forall J \in \mathbf{N}^+, \quad \forall j \in \llbracket 1, J \rrbracket, \quad \left( \mathbf{E} \left[ \sup_{t \in [0, T]} \left| X_t^{(j)} - \overline{X}_t^{(j)} \right|^p \right] \right)^{\frac{1}{p}} \leq C J^{-\min \left\{ \frac{1}{2}, \frac{q-p}{2p^2}, \frac{q-2p_{\mathcal{M}}}{2(2p_{\mathcal{M}})^2} \right\}}. \quad (2.20)$$

We summarize below the additional auxiliary results necessary for the proof, which are valid for an objective function  $f \in \mathcal{A}(s, \ell, \ell)$ . These results are stated precisely and proved in Section 3.

- **Wasserstein stability estimate for the weighted covariance.** We prove in Corollary 3.3 that, for all  $R > 0$  and for all  $p \geq 2p_{\mathcal{M}}(\ell, s)$ , there exists  $\tilde{L}_{\sqrt{C}}(R, p) > 0$  such that

$$\forall (\mu, \nu) \in \mathcal{P}_{p,R}(\mathbf{R}^d) \times \mathcal{P}_p(\mathbf{R}^d), \quad \left\| \sqrt{\mathcal{C}_\beta(\mu)} - \sqrt{\mathcal{C}_\beta(\nu)} \right\| \leq \tilde{L}_{\sqrt{C}} W_p(\mu, \nu).$$

- **Moment estimate for the particle system.** We prove in Lemma 3.5 that, under the assumptions of Theorem 2.12, there exists a constant  $\kappa > 0$  independent of  $J$  such that for all  $J \in \mathbf{N}^+$ ,

$$\mathbf{E} \left[ \sup_{t \in [0, T]} \left| X_t^{(j)} \right|^q \right] \vee \mathbf{E} \left[ \sup_{t \in [0, T]} \left| \mathcal{M}_\beta(\mu_t^J) \right|^q \right] \vee \mathbf{E} \left[ \sup_{t \in [0, T]} \left| \sqrt{\mathcal{C}_\beta(\mu_t^J)} \right|^q \right] \leq \kappa.$$

- **Convergence of the square root of the weighted covariance for i.i.d. samples.** We prove

in [Lemma 3.8](#) that, for any reals  $r > 2p > 0$  and any  $\mu \in \mathcal{P}_r(\mathbf{R}^d)$  such that  $\mathcal{C}_\beta(\mu) \succ 0$ , then there is  $C$  such that

$$\mathbf{E} \left\| \sqrt{\mathcal{C}_\beta(\bar{\mu}^J)} - \sqrt{\mathcal{C}_\beta(\mu)} \right\|^p \leqslant C J^{-\frac{p}{2}}, \quad \bar{\mu}^J := \frac{1}{J} \sum_{j=1}^J \delta_{\bar{X}^{(j)}}, \quad \{\bar{X}^{(j)}\}_{j \in \mathbf{N}} \stackrel{\text{i.i.d.}}{\sim} \mu.$$

- **No collapse in finite time.** In order to apply the third item with  $\mu = \bar{\rho}_t$ , we prove in [Proposition 3.9](#) that the assumption  $\mathcal{C}(\bar{\rho}_0) \succ 0$  implies that

$$\eta := \inf_{t \in [0, T]} \lambda_{\min}(\mathcal{C}_\beta(\bar{\rho}_t)) > 0.$$

### 3 Auxiliary results

This section is organized as follows. In [Subsection 3.1](#), we improve on existing Wasserstein stability estimates for the weighted mean and prove novel ones for the weighted covariance and the square root of the weighted covariance. In [Subsection 3.2](#), we present moment estimates for the empirical measures associated with the CBO (1.1) and CBS (1.4) dynamics. In [Subsection 3.3](#), we prove convergence estimates for the weighted moments when particle positions are drawn independently from a given probability distribution, using results from the statistics literature. Finally, in [Subsection 3.4](#), we prove that if  $\mathcal{C}_\beta(\bar{\rho}_0) \succ 0$ , then the covariance matrix  $\mathcal{C}_\beta(\bar{\rho}_t) \succ 0$  associated with the solution to the mean-field CBS dynamics (1.5) remains strictly positive definite for all times.

#### 3.1 Wasserstein stability estimates for the weighted moments

In this subsection, we generalize [[11](#), Lemma 3.2] to the more general conditions in [Assumption H](#) and, more importantly, to the weighted covariance and its square root.

**Proposition 3.1** (Basic stability estimates for the weighted mean and the weighted covariance). *Suppose that  $f \in \mathcal{A}(s, \ell, u)$  with  $s \geqslant 0$  and  $u \geqslant \ell \geqslant 0$ . Then:*

- (a) *For all  $R > 0$  and for all  $p \geqslant p_{\mathcal{M}}(s, \ell)$ , where  $p_{\mathcal{M}}(s, \ell)$  is given in (2.3), there exists  $L_{\mathcal{M}} = L_{\mathcal{M}}(p, R) > 0$  such that*

$$\forall (\mu, \nu) \in \mathcal{P}_{p, R}(\mathbf{R}^d) \times \mathcal{P}_{p, R}(\mathbf{R}^d), \quad |\mathcal{M}_\beta(\mu) - \mathcal{M}_\beta(\nu)| \leqslant L_{\mathcal{M}} W_p(\mu, \nu). \quad (3.1)$$

- (b) *For all  $R > 0$  and for all  $p \geqslant p_{\mathcal{C}}(s, \ell)$ , where  $p_{\mathcal{C}}(s, \ell)$  is given in (2.3), there exists  $L_{\mathcal{C}} = L_{\mathcal{C}}(p, R) > 0$  such that*

$$\forall (\mu, \nu) \in \mathcal{P}_{p, R}(\mathbf{R}^d) \times \mathcal{P}_{p, R}(\mathbf{R}^d), \quad \left\| \mathcal{C}_\beta(\mu) - \mathcal{C}_\beta(\nu) \right\|_{\mathbb{F}} \leqslant L_{\mathcal{C}} W_p(\mu, \nu). \quad (3.2)$$

*Proof.* The proof for both (a) and then (b) relies on the technical result [Lemma A.1](#) in the appendix.

**Proof of (a).** In the notation of [Lemma A.1](#) we can write

$$\mathcal{M}_\beta(\mu) = \frac{\mu[g]}{\mu[h]}, \quad g(x) := e^{-\beta f(x)} x, \quad h(x) := e^{-\beta f(x)}.$$

Condition (A.1) is satisfied for  $\varsigma := s + 1$  if  $\ell = 0$  or  $\varsigma := 0$  if  $\ell > 0$ . Therefore, the claim follows from [Lemma A.1](#).

**Proof of (b).** First observe that

$$\mathcal{C}_\beta(\mu) = \frac{\mu[g]}{\mu[h]} - \mathcal{M}_\beta(\mu) \otimes \mathcal{M}_\beta(\mu), \quad g(x) := e^{-\beta f(x)}(x \otimes x), \quad h(x) := e^{-\beta f(x)}.$$

By the triangle inequality, we have that

$$\|\mathcal{C}_\beta(\mu) - \mathcal{C}_\beta(\nu)\|_{\mathbb{F}} \leq \left\| \mathcal{M}_\beta(\mu) \otimes \mathcal{M}_\beta(\mu) - \mathcal{M}_\beta(\nu) \otimes \mathcal{M}_\beta(\nu) \right\|_{\mathbb{F}} + \left\| \frac{\mu[g]}{\mu[h]} - \frac{\nu[g]}{\nu[h]} \right\|_{\mathbb{F}}. \quad (3.3)$$

We shall use that for all  $(x, y) \in \mathbf{R}^d \times \mathbf{R}^d$ , it holds that

$$\begin{aligned} \|x \otimes x - y \otimes y\|_{\mathbb{F}} &= \left\| x(x-y)^{\top} + (x-y)y^{\top} \right\|_{\mathbb{F}} \\ &\leq \left\| x(x-y)^{\top} \right\|_{\mathbb{F}} + \left\| (x-y)y^{\top} \right\|_{\mathbb{F}} = |x| \cdot |x-y| + |y| \cdot |x-y| = (|x| + |y|)|x-y|. \end{aligned} \quad (3.4)$$

Using this inequality, that  $p_{\mathcal{C}} \geq p_{\mathcal{M}}$  and (a), we obtain for the first term in (3.3) that

$$\begin{aligned} \left\| \mathcal{M}_\beta(\mu) \otimes \mathcal{M}_\beta(\mu) - \mathcal{M}_\beta(\nu) \otimes \mathcal{M}_\beta(\nu) \right\|_{\mathbb{F}} &\leq \left( |\mathcal{M}_\beta(\mu)| + |\mathcal{M}_\beta(\nu)| \right) \cdot |\mathcal{M}_\beta(\mu) - \mathcal{M}_\beta(\nu)| \\ &\leq L_{\mathcal{M}} \left( |\mathcal{M}_\beta(\mu)| + |\mathcal{M}_\beta(\nu)| \right) W_p(\mu, \nu) \\ &\leq L_{\mathcal{M}}^2 \left( W_p(\mu, \delta_0) + W_p(\nu, \delta_0) \right) W_p(\mu, \nu) \leq 2L_{\mathcal{M}}^2 R W_p(\mu, \nu). \end{aligned}$$

In order to bound the second term in (3.3), we again use (3.4) to deduce that the function  $g$  satisfies the local Lipschitz estimate

$$\forall x, y \in \mathbf{R}^d, \quad |g(x) - g(y)| \leq L(1 + |x| + |y|)^{\varsigma} |x - y|, \quad \varsigma := \begin{cases} s + 2 & \text{if } \ell = 0, \\ 0 & \text{if } \ell > 0. \end{cases}$$

We deduce that the functions  $g$  and  $h$  satisfy the assumption (A.1) of Lemma A.1 with this  $\varsigma$ . Therefore, it holds that

$$\forall (\mu, \nu) \in \mathcal{P}_{p,R}(\mathbf{R}^d) \times \mathcal{P}_{p,R}(\mathbf{R}^d), \quad \left\| \frac{\mu[e^{-\beta f(x)} x \otimes x]}{\mu[e^{-\beta f(x)}]} - \frac{\nu[e^{-\beta f(x)} x \otimes x]}{\nu[e^{-\beta f(x)}]} \right\|_{\mathbb{F}} \leq C W_p(\mu, \nu),$$

which completes the proof.  $\square$

**Proposition 3.2** (Basic stability estimate for the square root covariance). *Suppose that  $f \in \mathcal{A}(s, \ell, u)$  with  $s \geq 0$  and  $u \geq \ell \geq 0$ . Then for all  $R > 0$  and all  $p \geq 2p_{\mathcal{M}}(s, \ell)$ , there exists  $L_{\sqrt{\mathcal{C}}} = L_{\sqrt{\mathcal{C}}}(p, R) > 0$  such that*

$$\forall (\mu, \nu) \in \mathcal{P}_{p,R}(\mathbf{R}^d) \times \mathcal{P}_{p,R}(\mathbf{R}^d), \quad \left\| \sqrt{\mathcal{C}_\beta(\mu)} - \sqrt{\mathcal{C}_\beta(\nu)} \right\|_{\mathbb{F}} \leq L_{\sqrt{\mathcal{C}}} W_p(\mu, \nu). \quad (3.5)$$

*Proof.* It is sufficient to check the claim for measures  $\mu$  and  $\nu$  of the form

$$\mu^J = \frac{1}{J} \sum_{j=1}^J \delta_{X^{(j)}}, \quad \nu^J = \frac{1}{J} \sum_{j=1}^J \delta_{Y^{(j)}}, \quad J \in \mathbf{N}^+. \quad (3.6)$$

Indeed, assume the claim holds for all such pairs of probability measures, fix  $R > 0$  and take any pair of probability measures  $(\mu, \nu) \in \mathcal{P}_{p,R} \times \mathcal{P}_{p,R}$ . By [52, Theorem 6.18], there exists a sequence  $\{(\mu^J, \nu^J)\}_{J \in \mathbf{N}^+}$  in  $\mathcal{P}_p \times \mathcal{P}_p$  such that  $W_p(\mu^J, \mu) \rightarrow 0$  and  $W_p(\nu^J, \nu) \rightarrow 0$  in the limit as  $J \rightarrow \infty$ . Since convergence in  $W_p$



implies convergence of the moments of order up to  $p$  by [52, Theorem 6.9], we can assume, discarding the beginning of the sequence if necessary, that  $(\mu^J, \nu^J) \in \mathcal{P}_{p,2R}$  for all  $J$ . Then

$$\begin{aligned} \left\| \sqrt{\mathcal{C}_\beta(\mu)} - \sqrt{\mathcal{C}_\beta(\nu)} \right\|_{\mathbb{F}} &\leq \left\| \sqrt{\mathcal{C}_\beta(\mu^J)} - \sqrt{\mathcal{C}_\beta(\nu^J)} \right\|_{\mathbb{F}} \\ &\quad + \left\| \sqrt{\mathcal{C}_\beta(\nu)} - \sqrt{\mathcal{C}_\beta(\nu^J)} \right\|_{\mathbb{F}} + \left\| \sqrt{\mathcal{C}_\beta(\mu)} - \sqrt{\mathcal{C}_\beta(\mu^J)} \right\|_{\mathbb{F}}. \end{aligned}$$

The first term is bounded from above by  $CW_p(\mu^J, \nu^J)$  by the base case, while the other two terms converge to 0 in the limit as  $J \rightarrow \infty$  by Proposition 3.1. Taking the limit  $J \rightarrow \infty$ , we deduce that

$$\left\| \sqrt{\mathcal{C}_\beta(\mu)} - \sqrt{\mathcal{C}_\beta(\nu)} \right\|_{\mathbb{F}} \leq CW_p(\mu, \nu).$$

**Proof of the statement for empirical measures.** Fix  $R > 0$ . By [53, p.5], the Wasserstein distance between empirical measures  $\mu^J$  and  $\nu^J$  of the form (3.6) is equal to

$$W_p(\mu^J, \nu^J) = \min_{\sigma \in \mathcal{S}_J} \left( \frac{1}{J} \sum_{j=1}^J |X^{(j)} - Y^{\sigma(j)}|^p \right)^{\frac{1}{p}},$$

where  $\mathcal{S}_J$  denotes the set of permutations in  $\{1, \dots, J\}$ . Thus, the claim will follow if we can prove that, for any pair  $(\mu^J, \nu^J) \in \mathcal{P}_{p,R}$  of the form (3.6), it holds that

$$\left\| \sqrt{\mathcal{C}_\beta(\mu^J)} - \sqrt{\mathcal{C}_\beta(\nu^J)} \right\|_{\mathbb{F}} \leq C \left( \frac{1}{J} \sum_{j=1}^J |X^{(j)} - Y^{(j)}|^p \right)^{\frac{1}{p}}. \quad (3.7)$$

We henceforth drop the superscript  $J$  in  $\mu^J, \nu^J$  for simplicity of notation, and write  $\mathbf{X} = (X^{(1)}, \dots, X^{(J)})$  and  $\mathbf{Y} = (Y^{(1)}, \dots, Y^{(J)})$ . First note that

$$\mathcal{C}_\beta(\mu) = \sum_{j=1}^J w_j(\mathbf{X}) \left( X^{(j)} - \mathcal{M}_\beta(\mu) \right) \otimes \left( X^{(j)} - \mathcal{M}_\beta(\mu) \right), \quad w_j(\mathbf{X}) := \frac{e^{-\beta f(X^{(j)})}}{\sum_{k=1}^J e^{-\beta f(X^{(k)})}}.$$

It is easily seen that  $\mathcal{C}_\beta(\mu) = M_{\mathbf{X}} M_{\mathbf{X}}^\top$ , where  $M_{\mathbf{X}}$  is the following  $\mathbf{R}^{d \times J}$  matrix

$$M_{\mathbf{X}} := \left( \sqrt{w_1(\mathbf{X})} (X^{(1)} - \mathcal{M}_\beta(\mu)) \quad \dots \quad \sqrt{w_J(\mathbf{X})} (X^{(J)} - \mathcal{M}_\beta(\mu)) \right).$$

Proceeding in the same manner, we construct a matrix  $M_{\mathbf{Y}} \in \mathbf{R}^{d \times J}$  such that  $\mathcal{C}_\beta(\nu) = M_{\mathbf{Y}} M_{\mathbf{Y}}^\top$ . A result by Araki and Yamagami [3], later generalized by Kittaneh [40] and Bhatia [7], states for any two matrices  $A$  and  $B$  with the same shape, it holds that

$$\left\| \sqrt{A^\top A} - \sqrt{B^\top B} \right\|_{\mathbb{F}} \leq \sqrt{2} \|A - B\|_{\mathbb{F}},$$

see also [6, Theorem VII.5.7] for a textbook presentation. This yields

$$\begin{aligned} & \left\| \sqrt{\mathcal{C}_\beta(\mu)} - \sqrt{\mathcal{C}_\beta(\nu)} \right\|_{\mathbb{F}} \\ & \leq \sqrt{2} \left( \sum_{j=1}^J \left| \sqrt{w_j(\mathbf{X})} (X^{(j)} - \mathcal{M}_\beta(\mu)) - \sqrt{w_j(\mathbf{Y})} (Y^{(j)} - \mathcal{M}_\beta(\nu)) \right|^2 \right)^{\frac{1}{2}} \\ & \leq \sqrt{2} \left( \sum_{j=1}^J \left| \sqrt{w_j(\mathbf{X})} X^{(j)} - \sqrt{w_j(\mathbf{Y})} Y^{(j)} \right|^2 \right)^{\frac{1}{2}} + \sqrt{2} \left( \sum_{j=1}^J \left| \sqrt{w_j(\mathbf{X})} \mathcal{M}_\beta(\mu) - \sqrt{w_j(\mathbf{Y})} \mathcal{M}_\beta(\nu) \right|^2 \right)^{\frac{1}{2}}, \end{aligned}$$

where we applied the triangle inequality for the Frobenius norm. Let  $\pi \in \Pi(\mu, \nu) \subset \mathcal{P}(\mathbf{R}^d \times \mathbf{R}^d)$  denote the probability measure

$$\pi = \frac{1}{J} \sum_{j=1}^J \delta_{(X^{(j)}, Y^{(j)})}$$

and set  $\gamma = \frac{\beta}{2}$ . We may rewrite the previous inequality as

$$\begin{aligned} \frac{1}{\sqrt{2}} \left\| \sqrt{\mathcal{C}_\beta(\mu)} - \sqrt{\mathcal{C}_\beta(\nu)} \right\|_{\mathbb{F}} & \leq \left( \iint_{\mathbf{R}^d \times \mathbf{R}^d} \left| \frac{e^{-\gamma f(x)} x}{\sqrt{\mu [e^{-\beta f}]}} - \frac{e^{-\gamma f(y)} y}{\sqrt{\nu [e^{-\beta f}]}} \right|^2 \pi(\mathrm{d}x \mathrm{d}y) \right)^{\frac{1}{2}} \\ & \quad + \left( \iint_{\mathbf{R}^d \times \mathbf{R}^d} \left| \frac{e^{-\gamma f(x)} \mathcal{M}_\beta(\mu)}{\sqrt{\mu [e^{-\beta f}]}} - \frac{e^{-\gamma f(y)} \mathcal{M}_\beta(\nu)}{\sqrt{\nu [e^{-\beta f}]}} \right|^2 \pi(\mathrm{d}x \mathrm{d}y) \right)^{\frac{1}{2}}. \end{aligned}$$

The conclusion then follows from [Lemma A.2](#) in the appendix:

- For the first term we let  $g(x) = e^{-\gamma f(x)} x$ ,  $h(x) = e^{-\beta f(x)}$ , and  $\mathcal{S}(\mu) = 1$ . The conditions of [Lemma A.2](#) are satisfied for  $\varsigma := s + 1$  if  $\ell = 0$  and  $\varsigma := 0$  if  $\ell > 0$ .
- For the second term let  $g(x) = e^{-\gamma f(x)}$ ,  $h(x) = e^{-\beta f(x)}$  and  $\mathcal{S}(\mu) = \mathcal{M}_\beta(\mu)$ . The conditions of [Lemma A.2](#) are satisfied for  $\varsigma := s$  if  $\ell = 0$  and  $\varsigma := 0$  if  $\ell > 0$ .

This completes the proof.  $\square$

The key for our mean-field proofs is the following improvement of [Propositions 3.1](#) and [3.2](#). Recall that

$$p_{\mathcal{M}} := p_{\mathcal{M}}(s, \ell) := \begin{cases} s + 2, & \text{if } \ell = 0, \\ 1, & \text{if } \ell > 0. \end{cases}$$

**Corollary 3.3** (Improved stability estimates). *Suppose that  $f \in \mathcal{A}(s, \ell, \ell)$  with  $s, \ell \geq 0$ . Then*

(a) *For all  $R > 0$  and for all  $p \geq p_{\mathcal{M}}$ , there exists  $\tilde{L}_{\mathcal{M}} = \tilde{L}_{\mathcal{M}}(R, p) > 0$  such that*

$$\forall (\mu, \nu) \in \mathcal{P}_{p, R}(\mathbf{R}^d) \times \mathcal{P}_p(\mathbf{R}^d), \quad |\mathcal{M}_\beta(\mu) - \mathcal{M}_\beta(\nu)| \leq \tilde{L}_{\mathcal{M}} W_p(\mu, \nu).$$

(b) *For all  $R > 0$  and  $p \geq 2p_{\mathcal{M}}$ , there exists  $\tilde{L}_{\mathcal{C}} = \tilde{L}_{\mathcal{C}}(R, p) > 0$  such that*

$$\forall (\mu, \nu) \in \mathcal{P}_{p, R}(\mathbf{R}^d) \times \mathcal{P}_p(\mathbf{R}^d), \quad \left\| \sqrt{\mathcal{C}_\beta(\mu)} - \sqrt{\mathcal{C}_\beta(\nu)} \right\|_{\mathbb{F}} \leq \tilde{L}_{\mathcal{C}} W_p(\mu, \nu).$$

**Remark 3.4.** *Observe that in contrast to [Propositions 3.1](#) and [3.2](#), no upper bounds on the  $p$ -th moments of  $\nu$  are needed in [Corollary 3.3](#), it is only required that they are finite.*

*Proof.* Let  $\mathcal{T} := \mathcal{M}_\beta$  or  $\mathcal{T} := \sqrt{\mathcal{C}_\beta}$ , and denote  $p_{\mathcal{T}} = p_{\mathcal{M}}$  if  $\mathcal{T} = \mathcal{M}_\beta$  or  $p_{\mathcal{T}} = 2p_{\mathcal{M}}$  if  $\mathcal{T} = \sqrt{\mathcal{C}_\beta}$ . Note that the following statements hold in either case.

- For all  $R > 0$  and  $p \geq p_{\mathcal{T}}$ , there exists  $L_{\mathcal{T}}(R) > 0$  such that

$$\forall (\mu, \nu) \in \mathcal{P}_{p,R}(\mathbf{R}^d) \times \mathcal{P}_{p,R}(\mathbf{R}^d), \quad \|\mathcal{T}(\mu) - \mathcal{T}(\nu)\| \leq L_{\mathcal{T}}(R) W_p(\mu, \nu). \quad (3.8)$$

Here,  $\|\cdot\|$  denotes either  $|\cdot|$  or  $\|\cdot\|_{\mathbb{F}}$ . Equation (3.8) follows from [Proposition 3.1](#) if  $\mathcal{T} = \mathcal{M}_\beta$  and [Proposition 3.2](#) if  $\mathcal{T} = \sqrt{\mathcal{C}_\beta}$ .

- For all  $p \geq p_{\mathcal{T}}$ , there is  $K > 0$  such that

$$\forall \mu \in \mathcal{P}_p(\mathbf{R}^d), \quad \|\mathcal{T}(\mu)\| \leq K W_p(\mu, \delta_0). \quad (3.9)$$

When  $\mathcal{T} = \mathcal{M}_\beta$ , this follows immediately from the inequality  $e^{-\beta f^*} \leq e^{-\beta f(\cdot)} \leq e^{-\beta f^*}$  if  $\ell = 0$  and from [Proposition A.3](#) if  $\ell > 0$ , where  $f^*$  the supremum of  $f$  over  $\mathbf{R}^d$ . When  $\mathcal{T} = \sqrt{\mathcal{C}_\beta}$ , the inequality (3.9) follows in the same way after noticing that

$$\left\| \sqrt{\mathcal{C}_\beta(\mu)} \right\|_{\mathbb{F}}^2 = \text{tr}(\mathcal{C}_\beta(\mu)) = \frac{\int_{\mathbf{R}^d} |x|^2 e^{-\beta f(x)} \mu(dx)}{\int_{\mathbf{R}^d} e^{-\beta f(x)} \mu(dx)} - \left| \mathcal{M}_\beta(\mu) \right|^2 \leq \frac{\int_{\mathbf{R}^d} |x|^2 e^{-\beta f(x)} \mu(dx)}{\int_{\mathbf{R}^d} e^{-\beta f(x)} \mu(dx)} \quad (3.10)$$

for all  $\mu \in \mathcal{P}_p(\mathbf{R}^d)$ .

Now, assume for contradiction that there exists  $R > 0$  and  $\{(\mu_i, \nu_i)\}_{i \in \mathbf{N}}$  in  $\mathcal{P}_{p,R}(\mathbf{R}^d) \times \mathcal{P}_p$  such that

$$\forall i \in \mathbf{N}, \quad \frac{\|\mathcal{T}(\mu_i) - \mathcal{T}(\nu_i)\|}{W_p(\mu_i, \nu_i)} \geq i. \quad (3.11)$$

In particular  $\nu_i[|x|^p] \rightarrow \infty$  in the limit as  $i \rightarrow \infty$ , otherwise the left-hand side of (3.11) would be bounded from above uniformly in  $i$  by (3.8). Therefore  $W_p(\nu_i, \delta_0) \rightarrow \infty$ . On the other hand, by the triangle inequality,

$$W_p(\mu_i, \nu_i) \geq W_p(\nu_i, \delta_0) - W_p(\mu_i, \delta_0) \geq W_p(\nu_i, \delta_0) - R.$$

Therefore, by (3.9), we deduce that

$$\limsup_{i \rightarrow \infty} \frac{\|\mathcal{T}(\mu_i) - \mathcal{T}(\nu_i)\|}{W_p(\mu_i, \nu_i)} \leq K \limsup_{i \rightarrow \infty} \frac{W_p(\nu_i, \delta_0) + R}{W_p(\nu_i, \delta_0) - R} = K,$$

which contradicts (3.11).  $\square$

### 3.2 Moment estimates for the empirical measures

**Lemma 3.5.** *Let  $f \in \mathcal{A}(s, \ell, \ell)$  with  $s, \ell \geq 0$ , and suppose that  $\bar{\rho}_0 \in \mathcal{P}_p(\mathbf{R}^d)$  for some  $p \geq 2$ .*

- (a) *Consider the particle system (1.1) with  $\bar{\rho}_0^{\otimes J}$ -distributed initial data, and let  $\mu_t^J$  denote the corresponding empirical measures. Then there exists a constant  $\kappa > 0$  independent of  $J$  such that for all  $J \in \mathbf{N}^+$*

$$\mathbf{E} \left[ \sup_{t \in [0, T]} |X_t^{(j)}|^p \right] \vee \mathbf{E} \left[ \sup_{t \in [0, T]} \mu_t^J[|x|^p] \right] \vee \mathbf{E} \left[ \sup_{t \in [0, T]} |\mathcal{M}_\beta(\mu_t^J)|^p \right] \leq \kappa.$$

(b) Consider the particle system (1.4) with  $\bar{\rho}_0^{\otimes J}$ -distributed initial data, and let  $\mu_t^J$  denote the corresponding empirical measures. Then there exists a constant  $\kappa > 0$  independent of  $J$  such that for all  $J \in \mathbf{N}^+$

$$\mathbf{E} \left[ \sup_{t \in [0, T]} |X_t^{(j)}|^p \right] \vee \mathbf{E} \left[ \sup_{t \in [0, T]} \mu_t^J[|x|^p] \right] \vee \mathbf{E} \left[ \sup_{t \in [0, T]} |\mathcal{M}_\beta(\mu_t^J)|^p \right] \vee \mathbf{E} \left[ \sup_{t \in [0, T]} \left\| \sqrt{\mathcal{C}_\beta(\mu_t^J)} \right\|_F^p \right] \leq \kappa.$$

*Proof.* The result can be proved using the same reasoning as in [11, Lemma 3.4] or [24, Lemma 15]. For convenience, we denote by  $C$  any constant depending only on  $(f, \beta, \lambda, p, T)$ . We first observe by (3.10) and Proposition A.3 that

$$\forall \mu \in \mathcal{P}_p(\mathbf{R}^d), \quad |\mathcal{M}_\beta(\mu)| \vee \left\| \sqrt{\mathcal{C}_\beta(\mu)} \right\|_F \leq C \left( \int_{\mathbf{R}^d} |x|^p d\mu \right)^{\frac{1}{p}}. \quad (3.12)$$

Writing the drift and diffusion coefficients for (a) and (b) as in (2.10) and (2.18), respectively, it holds in both cases that

$$\forall x \in \mathbf{R}^d, \quad \forall \mu \in \mathcal{P}(\mathbf{R}^d), \quad |b(x, \mu)| \vee \|\sigma(x, \mu)\|_F \leq C \left( |x|^p + \int_{\mathbf{R}^d} |y|^p \mu(dy) \right)^{\frac{1}{p}}.$$

Arguing in parallel for (a) and (b), we fix  $j \in \llbracket 1, J \rrbracket$  and apply the Burkholder–Davis–Gundy inequality to find for all  $t \in [0, T]$  that

$$\begin{aligned} \frac{1}{3^{p-1}} \mathbf{E} \left[ \sup_{s \in [0, t]} |X_s^{(j)}|^p \right] &\leq \mathbf{E} |X_0^{(j)}|^p + T^{p-1} \int_0^t \mathbf{E} |b(X_s^{(j)}, \mu_s^J)|^p ds \\ &\quad + C_{\text{BDG}} T^{\frac{p}{2}-1} \int_0^t \mathbf{E} \left\| \sigma(X_s^{(j)}, \mu_s^J) \right\|_F^p ds \\ &\leq \mathbf{E} |X_0^{(j)}|^p + C \int_0^t \mathbf{E} \left[ |X_s^{(j)}|^p + \int_{\mathbf{R}^d} |y|^p \mu_s^J(dy) \right] ds. \end{aligned}$$

Since the common law of  $(X^{(j)}, \dots, X^{(j)})$  is invariant under permutations of  $\llbracket 1, J \rrbracket$ , its marginals corresponding to the particles are all the same. Therefore, given that  $\mu_s^J = \frac{1}{J} \sum_{j=1}^J \delta_{X_s^{(j)}}$ , we find that

$$\mathbf{E} \left[ \int_{\mathbf{R}^d} |y|^p \mu_s^J(dy) \right] = \mathbf{E} |X_s^{(j)}|^p.$$

Consequently, it holds that

$$\mathbf{E} \left[ \sup_{s \in [0, t]} |X_s^{(j)}|^p \right] \leq C \mathbf{E} |X_0^{(j)}|^p + C \int_0^t \mathbf{E} \left[ \sup_{u \in [0, s]} |X_u^{(j)}|^p \right] ds.$$

Grönwall's inequality then gives that

$$\mathbf{E} \left[ \sup_{t \in [0, T]} |X_t^{(j)}|^p \right] \leq \kappa.$$

The second bound in (a) and in (b) follows from the inequality

$$\mathbf{E} \left[ \sup_{t \in [0, T]} \mu_t^J[|x|^p] \right] = \mathbf{E} \left[ \sup_{t \in [0, T]} \frac{1}{J} \sum_{j=1}^J |X_t^{(j)}|^p \right] \leq \mathbf{E} \left[ \frac{1}{J} \sum_{j=1}^J \sup_{t \in [0, T]} |X_t^{(j)}|^p \right] = \mathbf{E} \left[ \sup_{t \in [0, T]} |X_t^{(j)}|^p \right],$$

while the third bound in (a) as well as the third and fourth bound in (b) follow from (3.12).  $\square$

### 3.3 Convergence of weighted moments for i.i.d. samples

For the proof of the lemmas in this section, we will need [19, Theorem 1], which we recall for the reader's convenience in Theorem 3.6 below, slightly adapted to the i.i.d. setting. We also refer to [1, Theorem 2.3], where similar results for  $p = 1$  and  $p = 2$  are proved in the context of importance sampling, with explicit constant prefactors.

**Theorem 3.6** ( *$L^p$  convergence of weighted sums*). *Let  $(w_j, V_j)_{j \in \mathbf{N}}$  be i.i.d. random variables with values in  $\mathbf{R} \times \mathbf{R}^d$  or  $\mathbf{R} \times \mathbf{R}^{d \times d}$  with  $w_j > 0$  almost surely, and set*

$$\widehat{N}_J = \frac{1}{J} \sum_{j=1}^J w_j V_j, \quad \widehat{D}_J = \frac{1}{J} \sum_{j=1}^J w_j, \quad \widehat{R}_J = \frac{\widehat{N}_J}{\widehat{D}_J}.$$

Let also  $N = \mathbf{E}[\widehat{N}_J]$ ,  $D = \mathbf{E}[\widehat{D}_J]$ , and  $R = N/D$ . Let  $0 < p < q$  and assume that for some  $c > 0$ ,

$$\|w_1\|_{L^q(\Omega)} \leq c, \quad \|V_1\|_{L^r(\Omega)} \leq c, \quad \|w_1 V_1\|_{L^s(\Omega)} \leq c, \quad r := \frac{p(q+2)}{q-p}, \quad s := \frac{pq}{q-p}.$$

Then there is  $C$  depending only on  $(c, p, q, r, s)$  such that

$$\|\widehat{R}_J - R\|_{L^p(\Omega)} \leq C J^{-\frac{1}{2}}.$$

Using Theorem 3.6, we now prove the convergence of the weighted mean and weighted covariance for empirical measures formed from i.i.d. samples, in the limit of infinitely many samples.

**Lemma 3.7** (*Convergence of the weighted mean for i.i.d. samples*). *Suppose that  $f: \mathbf{R}^d \rightarrow \mathbf{R}$  is bounded from below and let  $0 < p < r$ . Then for all  $\mu \in \mathcal{P}_r(\mathbf{R}^d)$ , there is  $C$  depending only on  $(f, \beta, p, r)$  and the  $r$ -th moment of  $\mu$  such that for all  $J \in \mathbf{N}$*

$$\mathbf{E} |\mathcal{M}_\beta(\overline{\mu}^J) - \mathcal{M}_\beta(\mu)|^p \leq C J^{-\frac{p}{2}}, \quad \overline{\mu}^J := \frac{1}{J} \sum_{j=1}^J \delta_{\overline{X}^{(j)}}, \quad \{\overline{X}^{(j)}\}_{j \in \mathbf{N}} \stackrel{\text{i.i.d.}}{\sim} \mu.$$

*Proof of Lemma 3.7.* We apply Theorem 3.6 with  $q = \frac{p(r+2)}{r-p}$ , in which case  $\frac{p(q+2)}{q-p} = r$ . In our setting

$$(w_j, V_j) = \left( e^{-\beta f(\overline{X}^{(j)})}, \overline{X}^{(j)} e^{-\beta f(\overline{X}^{(j)})} \right).$$

Since  $w_j$  is bounded from above almost surely, it obviously holds that  $\|w_1\|_{L^q(\Omega)} < \infty$ . Since  $\mu \in \mathcal{P}_r(\mathbf{R}^d)$ , it also holds that  $\|V_1\|_{L^r(\Omega)} < \infty$  and  $\|w_1 V_1\|_{L^s(\Omega)} < \infty$  since  $s := \frac{pq}{q-p} < r$ . Consequently, the assumptions of Theorem 3.6 are satisfied, and we obtain the statement.  $\square$

**Lemma 3.8** (*Convergence of the weighted covariance for i.i.d. samples*). *Suppose that  $f: \mathbf{R}^d \rightarrow \mathbf{R}$  is bounded from below and let  $0 < 2p < r$ . Then for all  $\mu \in \mathcal{P}_r(\mathbf{R}^d)$  satisfying  $\mathcal{C}_\beta(\mu) \succcurlyeq \eta \text{Id} \succcurlyeq 0$ , there is  $C$  depending only on  $(f, \beta, p, r, \eta)$  and the  $r$ -th moment of  $\mu$  such that for all  $J \in \mathbf{N}^+$*

$$\mathbf{E} \left\| \sqrt{\mathcal{C}_\beta(\overline{\mu}^J)} - \sqrt{\mathcal{C}_\beta(\mu)} \right\|_{\mathbb{F}}^p \leq C J^{-\frac{p}{2}}, \quad \overline{\mu}^J := \frac{1}{J} \sum_{j=1}^J \delta_{X^{(j)}}, \quad \{X^{(j)}\}_{j \in \mathbf{N}} \stackrel{\text{i.i.d.}}{\sim} \mu.$$

*Proof.* By assumption it holds that  $\sqrt{\mathcal{C}_\beta(\overline{\mu}^J)} \succcurlyeq \sqrt{\eta} \text{Id}$ . By an inequality due to van Hemmen and Ando [51],

see also [6, Problem X.5.5], it holds that

$$\left\| \sqrt{\mathcal{C}_\beta(\bar{\mu}^J)} - \sqrt{\mathcal{C}_\beta(\mu)} \right\|_{\mathbb{F}} \leq \frac{1}{\eta} \left\| \mathcal{C}_\beta(\bar{\mu}^J) - \mathcal{C}_\beta(\mu) \right\|_{\mathbb{F}}.$$

Using the triangle inequality, we have that

$$\begin{aligned} \left\| \mathcal{C}_\beta(\bar{\mu}^J) - \mathcal{C}_\beta(\mu) \right\|_{\mathbb{F}}^p &= \left\| L_\beta \bar{\mu}^J [x \otimes x] - L_\beta \mu [x \otimes x] - \mathcal{M}_\beta(\bar{\mu}^J) \otimes \mathcal{M}_\beta(\bar{\mu}^J) + \mathcal{M}_\beta(\mu) \otimes \mathcal{M}_\beta(\mu) \right\|_{\mathbb{F}}^p \\ &\leq 2^{p-1} \left\| L_\beta \bar{\mu}^J [x \otimes x] - L_\beta \mu [x \otimes x] \right\|_{\mathbb{F}}^p \\ &\quad + 2^{p-1} \left\| \mathcal{M}_\beta(\bar{\mu}^J) \otimes \mathcal{M}_\beta(\bar{\mu}^J) - \mathcal{M}_\beta(\mu) \otimes \mathcal{M}_\beta(\mu) \right\|_{\mathbb{F}}^p. \end{aligned}$$

Convergence to zero of the expectation of the first term with rate  $J^{-\frac{p}{2}}$  follows from [Theorem 3.6](#), this time with  $q = \frac{p(r+4)}{r-2p}$ , in which case  $\frac{p(q+2)}{q-p} = \frac{r}{2}$ . For the second term, we use (3.4) to obtain that

$$\begin{aligned} &\mathbf{E} \left\| \mathcal{M}_\beta(\bar{\mu}^J) \otimes \mathcal{M}_\beta(\bar{\mu}^J) - \mathcal{M}_\beta(\mu) \otimes \mathcal{M}_\beta(\mu) \right\|_{\mathbb{F}}^p \\ &\leq \mathbf{E} \left[ (|\mathcal{M}_\beta(\bar{\mu}^J)| + |\mathcal{M}_\beta(\mu)|)^p |\mathcal{M}_\beta(\bar{\mu}^J) - \mathcal{M}_\beta(\mu)|^p \right] \\ &\leq \left( \mathbf{E} \left[ (|\mathcal{M}_\beta(\bar{\mu}^J)| + |\mathcal{M}_\beta(\mu)|)^{2p} \right] \mathbf{E} [|\mathcal{M}_\beta(\bar{\mu}^J) - \mathcal{M}_\beta(\mu)|^{2p}] \right)^{\frac{1}{2}} \\ &\leq C J^{-\frac{p}{2}} \left( \mathbf{E} \left[ (|\mathcal{M}_\beta(\bar{\mu}^J)| + |\mathcal{M}_\beta(\mu)|)^{2p} \right] \right)^{\frac{1}{2}}, \end{aligned}$$

where we used [Lemma 3.7](#). Since  $\mathcal{M}_\beta(\bar{\mu}^J)$  converges to  $\mathcal{M}_\beta(\mu)$  in  $L^{2p}(\Omega)$  again by [Lemma 3.7](#), this inequality enables to conclude the proof.  $\square$

### 3.4 No collapse in finite time for CBS

In order to deal with the matrix square root in the CBS dynamics, we prove that the weighted covariance stays positive definite for finite times.

**Proposition 3.9** (No collapse in finite time). *Let the cost function  $f: \mathbf{R}^d \rightarrow \mathbf{R}$  be such that the macroscopic McKean SDE for CBS (1.5) has a solution  $(\bar{X}_t)_{t \in [0, T]}$  over some finite time interval  $[0, T]$ . Then*

$$\forall t \in [0, T], \quad \mathcal{C}(\bar{\rho}_t) \succ \mathcal{C}(\bar{\rho}_0) e^{-2t}. \quad (3.13)$$

In particular, if  $\mathcal{C}(\bar{\rho}_0) \succ 0$  and if  $t \mapsto \mathcal{C}_\beta(\bar{\rho}_t)$  is continuous, then

$$\inf_{t \in [0, T]} \lambda_{\min}(\mathcal{C}(\bar{\rho}_t)) > 0 \quad \text{and} \quad \inf_{t \in [0, T]} \lambda_{\min}(\mathcal{C}_\beta(\bar{\rho}_t)) > 0. \quad (3.14)$$

*Proof.* The evolution of the covariance matrix is governed by [12, Equation (2.18)]

$$\partial_t (\mathcal{C}(\bar{\rho}_t)) = -2\mathcal{C}(\bar{\rho}_t) + 2\lambda^{-1}\mathcal{C}_\beta(\bar{\rho}_t).$$

For a given  $y \in \mathbf{R}^d$  such that  $|y| = 1$ , define  $g(t) := y^\top \mathcal{C}(\bar{\rho}_t) y$  for  $t \in [0, T]$ . Then

$$g'(t) = -2g(t) + 2\lambda^{-1} y^\top \mathcal{C}_\beta(\bar{\rho}_t) y \geq -2g(t),$$

since  $\mathcal{C}_\beta(\bar{\rho}_t)$  is positive semi-definite. Applying Grönwall's inequality to  $-g$ , we deduce that  $g(t) \geq g(0) e^{-2t}$ . This shows (3.13) and the first inequality in (3.14). In order to prove the second inequality in (3.14), assume that it was wrong. By continuity of  $t \mapsto \mathcal{C}_\beta(\bar{\rho}_t)$ , we have  $\lambda_{\min}(\mathcal{C}_\beta(\bar{\rho}_t)) = 0$  for some  $t \in [0, T]$ .

Then  $y^\top \mathcal{C}_\beta(\bar{\rho}_t)y = 0$  for some  $y \in \mathbf{R}^d$  such that  $|y| = 1$ . In particular

$$\int_{\mathbf{R}^d} \left| y^\top (x - \mathcal{M}_\beta(\bar{\rho}_t)) \right|^2 e^{-\beta f(x)} \bar{\rho}_t(dx) = 0.$$

Thus we obtain  $y^\top (x - \mathcal{M}_\beta(\bar{\rho}_t)) = 0$  for  $\bar{\rho}_t$ -almost every  $x \in \mathbf{R}^d$ , and so  $y^\top \mathcal{M}_\beta(\bar{\rho}_t) = y^\top \mathcal{M}(\bar{\rho}_t)$  by integrating against  $\bar{\rho}_t$ . But then,

$$y^\top \mathcal{C}(\bar{\rho}_t)y = \int_{\mathbf{R}^d} \left| y^\top (x - \mathcal{M}(\bar{\rho}_t)) \right|^2 \bar{\rho}_t(dx) = \int_{\mathbf{R}^d} \left| y^\top (x - \mathcal{M}_\beta(\bar{\rho}_t)) \right|^2 \bar{\rho}_t(dx) = 0.$$

which is a contradiction. Thus  $C_\beta(\bar{\rho}_t) \succ 0$  for all  $t \in [0, T]$ , which implies the second inequality in (3.14).  $\square$

## 4 Proof of the well-posedness results

### 4.1 Well-posedness for the microscopic models

For a collection of particles  $\mathbf{X} \in \mathbf{R}^{dJ}$ , the notation  $|\mathbf{X}|$  denotes the usual Euclidean norm; that is

$$|\mathbf{X}| := \sqrt{\sum_{j=1}^J |X^{(j)}|^2}.$$

For the proof of [Theorem 2.2](#) we need the following stability result, which only requires local Lipschitz continuity of  $f$ , in contrast to the results in [Subsection 3.1](#).

**Lemma 4.1.** *For all  $\mathbf{X} \in \mathbf{R}^{dJ}$  it holds that*

$$|\mathcal{M}_\beta(\mathbf{X})| \leq |\mathbf{X}| \quad \text{and} \quad \|\mathcal{C}_\beta(\mathbf{X})\| \leq |\mathbf{X}|^2. \quad (4.1)$$

Moreover, if  $f: \mathbf{R}^d \rightarrow \mathbf{R}$  is locally Lipschitz continuous, then for each  $R > 0$  and  $J \in \mathbf{N}^+$  there exists a constant  $C_{J,R} > 0$  such that

$$|\mathcal{M}_\beta(\mathbf{X}) - \mathcal{M}_\beta(\mathbf{Y})| \leq C_{J,R} |\mathbf{X} - \mathbf{Y}| \quad \text{and} \quad \left\| \sqrt{\mathcal{C}_\beta(\mathbf{X})} - \sqrt{\mathcal{C}_\beta(\mathbf{Y})} \right\|_F \leq C_{J,R} |\mathbf{X} - \mathbf{Y}| \quad (4.2)$$

for all  $\mathbf{X}, \mathbf{Y} \in \mathbf{R}^{dJ}$  satisfying  $|\mathbf{X}|, |\mathbf{Y}| \leq R$ .

*Proof.* The bound on  $\mathcal{M}_\beta(\mathbf{X})$  in (4.1) follows immediately from the definition: We have that

$$\left| \mathcal{M}_\beta(\mathbf{X}) \right|^2 = \left| \sum_{j=1}^J w_j(\mathbf{X}) X^{(j)} \right|^2 \leq \sum_{j=1}^J w_j(\mathbf{X}) \left| X^{(j)} \right|^2 \leq \sum_{j=1}^J \left| X^{(j)} \right|^2 = |\mathbf{X}|^2,$$

where we used convexity of the square function and the fact that the weights, given hereafter, are bounded from above by 1:

$$w_j(\mathbf{X}) := \frac{e^{-\beta f(X^{(j)})}}{\sum_{k=1}^J e^{-\beta f(X^{(k)})}}.$$

Similarly, for the weighted covariance, it holds by the triangle inequality that

$$\begin{aligned} \|\mathcal{C}_\beta(\mathbf{X})\|_{\mathbb{F}} &= \left\| \sum_{j=1}^J w_j(\mathbf{X}) \left( X^{(j)} \otimes X^{(j)} \right) - \mathcal{M}_\beta(\mathbf{X}) \otimes \mathcal{M}_\beta(\mathbf{X}) \right\|_{\mathbb{F}} \\ &\leq \left\| \sum_{j=1}^J w_j(\mathbf{X}) \left( X^{(j)} \otimes X^{(j)} \right) \right\|_{\mathbb{F}} \leq \sum_{j=1}^J w_j(\mathbf{X}) \|X^{(j)} \otimes X^{(j)}\|_{\mathbb{F}} = \sum_{j=1}^J w_j(\mathbf{X}) |X^{(j)}|^2 \leq |\mathbf{X}|^2. \end{aligned}$$

The local Lipschitz estimate on  $\mathcal{M}_\beta(\mathbf{X})$  in (4.2) is [11, Lemma 2.1]. In order to prove the local Lipschitz continuity of  $\mathcal{C}_\beta(\mathbf{X})$ , fix  $R > 0$  and  $J > 0$ . As in the proof of Proposition 3.2 we obtain

$$\begin{aligned} \left\| \sqrt{\mathcal{C}_\beta(\mathbf{X})} - \sqrt{\mathcal{C}_\beta(\mathbf{Y})} \right\|_{\mathbb{F}} &\leq \sqrt{2} \left( \sum_{j=1}^J \left| \sqrt{w_j(\mathbf{X})} \left( X^{(j)} - \mathcal{M}_\beta(\mathbf{X}) \right) - \sqrt{w_j(\mathbf{Y})} \left( Y^{(j)} - \mathcal{M}_\beta(\mathbf{Y}) \right) \right|^2 \right)^{\frac{1}{2}} \\ &\leq \sqrt{2} \sum_{j=1}^J \left| \sqrt{w_j(\mathbf{X})} \left( X^{(j)} - \mathcal{M}_\beta(\mathbf{X}) \right) - \sqrt{w_j(\mathbf{Y})} \left( Y^{(j)} - \mathcal{M}_\beta(\mathbf{Y}) \right) \right|. \end{aligned}$$

Using the triangle inequality, we have that

$$\begin{aligned} \frac{1}{\sqrt{2}} \left\| \sqrt{\mathcal{C}_\beta(\mathbf{X})} - \sqrt{\mathcal{C}_\beta(\mathbf{Y})} \right\|_{\mathbb{F}} &\leq \sum_{j=1}^J \sqrt{w_j(\mathbf{X})} \left| X^{(j)} - \mathcal{M}_\beta(\mathbf{X}) - Y^{(j)} + \mathcal{M}_\beta(\mathbf{Y}) \right| \\ &\quad + \sum_{j=1}^J \left| \left( \sqrt{w_j(\mathbf{X})} - \sqrt{w_j(\mathbf{Y})} \right) \left( Y^{(j)} - \mathcal{M}_\beta(\mathbf{Y}) \right) \right|. \end{aligned}$$

Since  $\mathbf{X} \mapsto \mathcal{M}_\beta(\mathbf{X})$  is a locally Lipschitz continuous function by [11, Lemma 2.1], it suffices to prove that  $\mathbf{X} \mapsto \sqrt{w_j(\mathbf{X})}$  is locally Lipschitz as well. To this end, observe that

$$\sqrt{w_j(\mathbf{X})} - \sqrt{w_j(\mathbf{Y})} = \frac{w_j(\mathbf{X}) - w_j(\mathbf{Y})}{\sqrt{w_j(\mathbf{X})} + \sqrt{w_j(\mathbf{Y})}}.$$

The denominator can be controlled by noting that

$$\forall |\mathbf{X}| \leq R, \quad w_j(\mathbf{X}) = \frac{e^{-\beta f(X^{(j)})}}{\sum_{k=1}^J e^{-\beta f(X^{(k)})}} \geq \frac{1}{J} \exp \left( -\beta \left( \max_{|x| \leq R} f(x) - \min_{|x| \leq R} f(x) \right) \right).$$

Therefore, there exists a constant  $c_{R,J} > 0$  such that  $\sqrt{w_j(\mathbf{X})} \geq c_{R,J}$  for all  $|\mathbf{X}| \leq R$ . Since  $\mathbf{X} \mapsto w_j(\mathbf{X})$  is locally Lipschitz continuous as a composition of locally Lipschitz continuous functions, the claim follows.  $\square$

*Proof of Theorem 2.2.* We prove only the claim for the CBS dynamics (1.4), since the proof for CBO is similar. We write the system (1.4) in the form

$$d\mathbf{X}_t = F(\mathbf{X}_t) dt + G(\mathbf{X}_t) d\mathbf{W}_t$$

where  $\mathbf{X}_t := (X_t^{(1)}, \dots, X_t^{(J)})$ . The drift and diffusion coefficients are given by

$$F(\mathbf{X}_t) := - \begin{pmatrix} X_t^{(1)} - \mathcal{M}_\beta(\mu_t^J) \\ \vdots \\ X_t^{(J)} - \mathcal{M}_\beta(\mu_t^J) \end{pmatrix} \in \mathbf{R}^{dJ},$$



and

$$G(\mathbf{X}_t) := \sqrt{2\lambda^{-1}} \operatorname{diag} \left( \sqrt{\mathcal{C}_\beta(\mu_t^J)}, \dots, \sqrt{\mathcal{C}_\beta(\mu_t^J)} \right) \in \mathbf{R}^{dJ \times dJ}.$$

By [Lemma 4.1](#), the functions  $F$  and  $G$  have sublinear growth and are locally Lipschitz continuous. Thus, by the non-explosion criterion from stochastic Lyapunov theory [[39](#), Theorem 3.5], it suffices to find a function  $\phi \in C^2(\mathbf{R}^{dJ}, [0, \infty))$  such that

- The function  $\phi$  is coercive:  $\phi(\mathbf{X}) \rightarrow \infty$  as  $|\mathbf{X}| \rightarrow \infty$  and
- There exists  $c > 0$  such that

$$\forall \mathbf{X} \in \mathbf{R}^{dJ}, \quad \mathcal{L}\phi(\mathbf{X}) := F(\mathbf{X}) \cdot \nabla \phi(\mathbf{X}) + \frac{1}{2} \operatorname{tr} \left( G^\top(\mathbf{X}) \mathbf{D}^2 \phi(\mathbf{X}) G(\mathbf{X}) \right) \leq c\phi(\mathbf{X}).$$

Take  $\phi(\mathbf{X}) := \frac{1}{2} |\mathbf{X}|^2$ . Then

$$\begin{aligned} \mathcal{L}\phi(\mathbf{X}) &= - \sum_{j=1}^J \left( X^{(j)} - \mathcal{M}_\beta(\mu_t^J) \right) \cdot X^{(j)} + J\lambda^{-1} \operatorname{tr} \left( \mathcal{C}_\beta(\mu_t^J) \right) \\ &= - \sum_{j=1}^J \left( X^{(j)} - \mathcal{M}_\beta(\mu_t^J) \right) \cdot X^{(j)} + \lambda^{-1} \sum_{j=1}^J w_j(\mathbf{X}) \left| X^{(j)} - \mathcal{M}_\beta(\mu_t^J) \right|^2. \end{aligned}$$

From [\(4.1\)](#) it follows that this is a valid choice, concluding the proof.  $\square$

## 4.2 Well-posedness for the mean-field SDEs

We adapt the argument from [[11](#)], using the more general [Assumption H](#) and the improved stability estimates from [Subsection 3.1](#). We first prove [Theorem 2.4](#), and then comment on the necessary changes for proving [Theorem 2.3](#) in [Remark 4.3](#). In the following, we endow the vector space  $\mathcal{X} := \mathbf{R}^d \times \mathbf{R}^{d \times d}$  with the norm

$$\|(m, \Gamma)\| := \left( |m|^2 + \|\Gamma\|_{\mathbf{F}}^2 \right)^{\frac{1}{2}}.$$

The Hölder space  $C^{0, \frac{1}{2}}([0, T], \mathcal{X})$  is given by

$$C^{0, \frac{1}{2}}([0, T], \mathcal{X}) := \left\{ \varphi: [0, T] \rightarrow \mathcal{X} : \|\varphi\|_{C^{0, \frac{1}{2}}} < \infty \right\},$$

where

$$\|\varphi\|_{C^{0, \frac{1}{2}}} := \sup_{t \in [0, T]} \|\varphi(t)\| + [\varphi]_{C^{0, \frac{1}{2}}} < \infty \quad \text{and} \quad [\varphi]_{C^{0, \frac{1}{2}}} := \sup_{0 \leq s < t \leq T} \frac{\|\varphi(t) - \varphi(s)\|}{|t - s|^{\frac{1}{2}}}.$$

In order to ensure that certain matrices are positive semidefinite, we use the map  $\mathcal{R}: \mathbf{R}^{d \times d} \rightarrow \mathbf{R}^{d \times d}$  defined by  $\mathcal{R}(\Gamma) = \sqrt{\Gamma \Gamma^\top}$ . We will make use of the following auxiliary lemma.

**Lemma 4.2.** *For  $f$  and  $\bar{\rho}_0$  as in [Theorem 2.4](#), fix  $(m, \Gamma) \in C([0, T], \mathcal{X})$  and  $y_0 \sim \bar{\rho}_0$ . Then the stochastic differential equation*

$$dY_t = -(Y_t - m_t) dt + \sqrt{2\lambda^{-1} \mathcal{R}(\Gamma_t)} dW_t, \quad Y_0 = y_0, \quad (4.3)$$

*admits a unique continuous strong solution  $Y \in C([0, T], \mathbf{R}^d)$ . Moreover, denoting by  $\rho_t := \operatorname{Law}(Y_t) \in \mathcal{P}(\mathbf{R}^d)$  the pointwise law of  $Y_t$ , the functions  $t \mapsto \mathcal{M}_\beta(\rho_t)$  and  $t \mapsto \mathcal{C}_\beta(\rho_t)$  belong to  $C([0, T], \mathbf{R}^d)$  and  $C([0, T], \mathbf{R}^{d \times d})$*

respectively.

*Proof.* The existence of a unique continuous strong solution  $Y \in C([0, T], \mathbf{R}^d)$  to (4.3) follows from classical theory of stochastic differential equations [45, Theorem 5.2.1]. It remains to prove the continuity statement for the weighted mean and covariance. For any  $0 \leq r \leq t \leq T$ , it holds by the Burkholder–Davis–Gundy inequality that

$$\begin{aligned} \mathbf{E} \left[ \sup_{u \in [r, t]} |Y_u - Y_r|^p \right] &\leq (2(t-r))^{p-1} \int_r^t \mathbf{E} |Y_u - m_u|^p \, du \\ &\quad + C_{\text{BDG}} 2^{p-1} (2\lambda^{-1})^{\frac{p}{2}} (t-r)^{\frac{p}{2}-1} \int_r^t \left\| \sqrt{\mathcal{R}(\Gamma(u))} \right\|_{\mathbb{F}}^p \, du. \end{aligned} \quad (4.4)$$

Taking  $r = 0$  in this inequality and rearranging, we obtain that

$$\mathbf{E} \left[ \sup_{u \in [0, t]} |Y_u|^p \right] \leq C \left( \mathbf{E} |Y_0|^p + \|m\|_{L^\infty([0, T])}^p + \|\Gamma\|_{L^\infty([0, T])}^{\frac{p}{2}} + \int_0^t \mathbf{E} |Y_u|^p \, du \right),$$

for a constant  $C$  depending only on  $(T, \lambda, p)$ . It follows from this inequality, by Grönwall's lemma, that

$$\mathbf{E} \left[ \sup_{t \in [0, T]} |Y_t|^p \right] \leq C \left( \mathbf{E} |Y_0|^p + \|m\|_{L^\infty([0, T])}^p + \|\Gamma\|_{L^\infty([0, T])}^{\frac{p}{2}} \right) e^{CT} < \infty. \quad (4.5)$$

Therefore, by dominated convergence, we have for any  $t \in [0, T]$  that

$$\mathcal{M}_\beta(\rho_r) = \frac{\mathbf{E} [Y_r e^{-\beta f(Y_r)}]}{\mathbf{E} [e^{-\beta f(Y_r)}]} \xrightarrow{r \rightarrow t} \frac{\mathbf{E} [Y_t e^{-\beta f(Y_t)}]}{\mathbf{E} [e^{-\beta f(Y_t)}]} = \mathcal{M}_\beta(\rho_t),$$

implying that the function  $t \mapsto \mathcal{M}_\beta(\rho_t)$  belongs to  $C([0, T], \mathbf{R}^d)$ . An analogous reasoning shows that the function  $t \mapsto \mathcal{C}_\beta(\rho_t)$  belongs to  $C([0, T], \mathbf{R}^{d \times d})$ .  $\square$

*Proof of Theorem 2.4.* Similarly to the proof of [11, Theorem 3.1], we first construct a map

$$\mathcal{T}: C([0, T], \mathcal{X}) \rightarrow C([0, T], \mathcal{X})$$

whose fixed points correspond to solutions of (1.5).

**Step 1: Constructing the map  $\mathcal{T}$ .** Consider the map

$$\begin{aligned} \mathcal{T}: C([0, T], \mathcal{X}) &\rightarrow C([0, T], \mathcal{X}) \\ (m, \Gamma) &\mapsto (\mathcal{M}_\beta(\rho_t), \mathcal{C}_\beta(\rho_t)), \end{aligned}$$

where  $\rho_t := \text{Law}(Y_t)$  and  $(Y_t)_{t \in [0, T]}$  is the unique solution to (4.3). Thanks to Lemma 4.2, the map  $\mathcal{T}$  is well-defined. Fixed points of  $\mathcal{T}$  correspond to solutions of the McKean-Vlasov SDE (1.5). The existence of fixed points follows from applying the Leray-Schauder Fixed Point Theorem [29, Chapter 11] in the space  $C([0, T], \mathcal{X})$ , once we have proved that  $\mathcal{T}$  is compact and that the following set is bounded:

$$\left\{ (m, \Gamma) \in C([0, T], \mathcal{X}) : \exists \xi \in [0, 1] \text{ such that } (m, \Gamma) = \xi \mathcal{T}((m, \Gamma)) \right\}. \quad (4.6)$$

**Step 2: Showing that  $\mathcal{T}$  is compact.** In order to prove that  $\mathcal{T}$  is a compact operator, fix  $R > 0$  and consider the ball

$$B_R := \left\{ (m, \Gamma) \in C([0, T], \mathcal{X}) : \|(m, \Gamma)\|_{L^\infty([0, T])} \leq R \right\}.$$

By the Arzelà–Ascoli theorem, we have a compact embedding

$$C^{0, \frac{1}{2}}([0, T], \mathcal{X}) \hookrightarrow C([0, T], \mathcal{X}).$$

Thus, it suffices to show that  $\mathcal{T}(B_R)$  is bounded in  $C^{0, \frac{1}{2}}([0, T], \mathcal{X})$ . To this end, take  $(m, \Gamma) \in B_R$  and consider the corresponding solution  $Y_t$  of (4.3) with pointwise law  $\rho_t$ , and let  $\varphi = \mathcal{T}(m, \Gamma)$ . We first observe that

$$\|\varphi(0)\|^2 = |\mathcal{M}_\beta(\bar{\rho}_0)|^2 + \|\mathcal{C}_\beta(\bar{\rho}_0)\|_{\mathbb{F}}^2 < \infty.$$

The right-hand side is independent of  $m$  and  $\Gamma$ , and it is indeed finite by [Proposition A.3](#). Given that

$$\|\varphi\|_{C^{0, \frac{1}{2}}} \leq \|\varphi(0)\| + [\varphi]_{C^{0, \frac{1}{2}}}(\sqrt{T} + 1),$$

it only remains to show that  $[\varphi]_{C^{0, \frac{1}{2}}}$  is bounded by a finite constant depending on  $(m, \Gamma)$  only through  $R$ . In view of (4.4) and (4.5), there is a constant  $L = L(R)$  such that

$$\mathbf{E} |Y_t - Y_r|^p \leq L|t - r|^{\frac{p}{2}}.$$

It follows that  $W_p(\rho_t, \rho_r) \leq L^{\frac{1}{p}}|t - r|^{\frac{1}{2}}$ , which by [Proposition 3.1](#) implies the Hölder continuity with exponent  $\frac{1}{2}$  of the functions  $t \mapsto \mathcal{M}_\beta(\rho_t)$  and  $t \mapsto \mathcal{C}_\beta(\rho_t)$ . This completes the proof that  $\mathcal{T}$  is compact.

**Step 3: Showing that the set (4.6) is bounded.** To this end, let  $(m, \Gamma) \in C([0, T], \mathcal{X})$  be such that

$$(m, \Gamma) = \xi \mathcal{T}(m, \Gamma) \tag{4.7}$$

for some  $\xi \in [0, 1]$ , and let  $(Y_t)$  denote the corresponding solution to (4.3). By (4.7), the stochastic process  $(Y_t)$  is also a solution to

$$dY_t = -\left(Y_t - \xi \mathcal{M}_\beta(\rho_t)\right) dt + \sqrt{2\lambda^{-1}\xi \mathcal{C}_\beta(\rho_t)} dW_t, \quad \rho_t := \text{Law}(Y_t).$$

But then, reasoning as in [Subsection 3.2](#) using a Grönwall-type argument and [Proposition A.3](#) with  $u = \ell$ , we find that

$$\mathbf{E} \left[ \sup_{t \in [0, T]} |Y_t|^p \right] \leq C \mathbf{E} |Y_0|^p = C \int_{\mathbf{R}^d} |x|^p \bar{\rho}_0(dx).$$

By [Proposition A.3](#), it follows that  $\mathcal{M}_\beta(\rho_t)$  and  $\mathcal{C}_\beta(\rho_t)$  can be bounded uniformly in  $[0, T]$  in terms of  $\bar{\rho}_0$ , implying that the set (4.6) is indeed bounded. Thus, we established the existence of a fixed point of  $\mathcal{T}$ , and we obtain as a byproduct the moment bounds in (2.6).

**Step 4: Showing uniqueness.** For uniqueness, let  $(m_t, \Gamma_t)$  and  $(\hat{m}_t, \hat{\Gamma}_t)$  be two fixed points of  $\mathcal{T}$  with corresponding solutions  $Y_t, \hat{Y}_t \in C([0, T], \mathcal{X})$  of (4.3) and let  $z_t := Y_t - \hat{Y}_t$ . By definition of  $\mathcal{T}$  and since  $\mathcal{C}_\beta(\mu)$  is symmetric and positive semidefinite for all  $\mu \in \mathcal{P}_2(\mathbf{R}^d)$ , we have  $\mathcal{R}(\Gamma_t) = \Gamma_t$  and  $\mathcal{R}(\hat{\Gamma}_t) = \hat{\Gamma}_t$  for all  $t \in [0, T]$ . Let  $\rho_t, \hat{\rho}_t \in \mathcal{P}(\mathbf{R}^d)$  denote the marginal laws of  $Y_t$  and  $\hat{Y}_t$ , respectively. By the Burkholder–Davis–Gundy inequality, we have for all  $t \in [0, T]$

$$\begin{aligned} \frac{1}{2^{p-1}} \mathbf{E} \left[ \sup_{r \in [0, t]} |z_r|^p \right] &\leq T^{p-1} \int_0^t \mathbf{E} \left| Y_r - \hat{Y}_r - \mathcal{M}_\beta(\rho_r) + \mathcal{M}_\beta(\hat{\rho}_r) \right|^p dr \\ &\quad + C_{\text{BDG}} T^{\frac{p}{2}-1} (2\lambda^{-1})^{\frac{p}{2}} \int_0^t \left\| \sqrt{\mathcal{C}_\beta(\rho_r)} - \sqrt{\mathcal{C}_\beta(\hat{\rho}_r)} \right\|_{\mathbb{F}}^p dr. \end{aligned}$$

By [Proposition 3.9](#), we have that

$$\eta := \inf \left\{ \lambda_{\min} \left( \mathcal{C}_\beta(\rho_t) \right) : t \in [0, T] \right\} > 0.$$

By virtue of the van Hemmen–Ando inequality [[51](#), Proposition 3.2], it holds for any unitarily invariant norm, thus in particular for the Frobenius norm, that

$$\left\| \sqrt{\mathcal{C}_\beta(\rho_r)} - \sqrt{\mathcal{C}_\beta(\widehat{\rho}_r)} \right\|_{\mathbb{F}} \leq \eta^{-\frac{1}{2}} \|\mathcal{C}_\beta(\rho_r) - \mathcal{C}_\beta(\widehat{\rho}_r)\|_{\mathbb{F}}.$$

Using [Proposition 3.1](#) and noting that  $W_p(\rho_r, \widehat{\rho}_r)^p \leq \mathbf{E}|z_r|^p$  by definition of the Wasserstein distance, we conclude that

$$\mathbf{E} \left[ \sup_{r \in [0, t]} |z_r|^p \right] \leq C \int_0^t \mathbf{E} \left[ \sup_{u \in [0, r]} |z_u|^p \right] dr.$$

By Grönwall’s lemma, and given that  $z_0 = 0$ , it follows that

$$\mathbf{E} \left[ \sup_{t \in [0, T]} |z_t|^p \right] = 0,$$

which concludes the proof of uniqueness.

**Step 5: Showing that  $\bar{\rho}_t$  solves the nonlocal Fokker–Planck equation (2.7).** This part of the proof is standard, and we present only a formal derivation. Let  $(m, \Gamma) \in C([0, T], \mathcal{X})$  denote the fixed point of  $\mathcal{T}$ , in which case  $\mathcal{R}(\Gamma_t) = \Gamma_t$ , and let  $\bar{Y}_t$  denote the corresponding solution to (4.3). Denote also by  $\bar{\rho}_t$  the law of  $\bar{Y}_t$ . For a smooth function  $\phi: \mathbf{R}^d \rightarrow \mathbf{R}$  with compact support, it holds by Itô’s lemma that

$$\mathbf{E}\phi(\bar{Y}_t) - \mathbf{E}\phi(\bar{Y}_r) = \mathbf{E} \left[ \int_r^t \mathcal{L}_u \phi(\bar{Y}_u) du \right], \quad (4.8)$$

where  $\mathcal{L}_u \phi(y) := -(y - \mathcal{M}_\beta(\bar{\rho}_u)) \cdot \nabla \phi(y) + \lambda^{-1} \mathcal{C}_\beta(\bar{\rho}_u) : D^2 \phi(y)$ . It follows from (4.8) that  $\bar{\rho}_t$  satisfies, for all smooth compactly supported  $\phi: \mathbf{R}^d \rightarrow \mathbf{R}$ ,

$$\frac{d}{dt} \int_{\mathbf{R}^d} \phi(x) \bar{\rho}_t(dx) = \int_{\mathbf{R}^d} \mathcal{L}_t \phi \bar{\rho}_t(dx), \quad (4.9)$$

which is indeed a weak formulation of the Fokker–Planck equation (2.5). For rigorous uniqueness and existence results for the Fokker–Planck equation (4.9) as well as a more precise discussion of its connection with (4.3), see for example [[26](#)].  $\square$

**Remark 4.3** (Proof of [Theorem 2.3](#)). *Observe that the drift coefficients of CBO and CBS are identical and that the diffusion coefficient of CBO is a globally Lipschitz continuous function of its drift coefficient. By, among other things, replacing  $\sqrt{2\lambda^{-1}\mathcal{R}(\Gamma_t)}$  in (4.3) by  $S(\bar{X}_t - \mathcal{M}_\beta(\bar{\rho}_t))$ , the above proof can be easily adapted to provide a proof for [Theorem 2.3](#). See also [[11](#)], where a weaker version of [Theorem 2.3](#) was already proven in the case where  $p = 4$  and  $f \in \mathcal{A}(1, 0, 0) \cup \mathcal{A}(1, 2, 2)$ .*

**Remark 4.4.** *Note that in case of  $p \geq 2p_{\mathcal{M}}$ , we can drop the requirement of  $\mathcal{C}(\bar{\rho}_0)$  to be positive definite by using [Proposition 3.2](#) instead of [Proposition 3.1](#).*

## 5 Extensions and perspectives for future work

This section is organized as follows. First, in [Subsection 5.1](#), we present an alternative approach based on stopping times for proving our main results [Theorems 2.6](#) and [2.12](#). Then, in [Subsection 5.2](#), we show how the stopping time approach presented in [Subsection 5.1](#) can be employed in order to obtain a mean-field result in the setting where  $f \in \mathcal{A}(s, \ell, u)$  with  $u > \ell > 0$ , under which the improved stability estimates in [Corollary 3.3](#) are no longer valid.

### 5.1 Proving mean-field results using stopping times

In this section, we present an alternative approach for proving [Theorem 2.6](#) and [Theorem 2.12](#). This approach is based on stopping times, and inspired by a method employed in [\[34\]](#) to handle local Lipschitz continuity issues in the context of numerical analysis for the Euler–Maruyama method. A similar approach was employed in [\[37\]](#) to obtain a non-quantitative mean-field estimate, and more recently in [\[50\]](#) to prove sharp propagation of chaos estimates for an ensemble sampler related to the ensemble Kalman filter [\[27, 28\]](#). Before presenting the proof, we briefly discuss the main differences between this approach and the indicator set approach presented in [Section 2](#).

- The main advantage of the approach used in [Section 2](#) is that the event  $\Omega_{J,t}$  for which our stability estimates do not apply can be time-dependent. For the proof of [Theorem 2.6](#), this flexibility is in fact not necessary, as we could have employed instead of  $\Omega_{J,t}$  in [\(2.16\)](#) the event

$$\Omega_J := \left\{ \omega \in \Omega : \frac{1}{J} \sum_{j=1}^J \sup_{t \in [0, T]} |\bar{X}_t^{(j)}(\omega)|^p \geq R \right\}. \quad (5.1)$$

In certain contexts, however, being able to exclude a time-dependent event could be very useful.

- The stopping time approach presented hereafter does not enjoy this flexibility, but is convenient in settings where moments of both  $\mu^J$  and  $\bar{\mu}^J$  need to be controlled in order to apply the stability estimates obtained in [Proposition 3.1](#) for the weighted moments. This is illustrated in [Subsection 5.2](#).

We illustrate the stopping times approach only for [Theorem 2.12](#), noting that [Theorem 2.6](#) can be proved in exactly the same manner.

*Alternative proof of [Theorem 2.12](#).* Let  $R > 0$  be such that the stopping time

$$\theta_J(R) := \inf \left\{ t \geq 0 : \frac{1}{J} \sum_{j=1}^J |X_t^{(j)}|^p \geq R \right\}, \quad (5.2)$$

satisfies  $\mathbf{P}[\theta_J \leq T] = \mathcal{O}(J^{-2})$ . Note that the event  $\{\theta_J \leq T\}$  coincides with the event  $\Omega_J$  given in [\(5.1\)](#). From Hölder’s inequality it follows that

$$\begin{aligned} & \frac{1}{2^{p-1}} \mathbf{E} \left[ \sup_{t \in [0, T]} |X_t^{(j)} - \bar{X}_t^{(j)}|^p \right] \\ &= \mathbf{E} \left[ \sup_{t \in [0, T]} |X_t^{(j)} - \bar{X}_t^{(j)}|^p \mathbf{1}_{\theta_J(R) > T} \right] + \mathbf{E} \left[ \sup_{t \in [0, T]} |X_t^{(j)} - \bar{X}_t^{(j)}|^p \mathbf{1}_{\theta_J(R) \leq T} \right] \\ &\leq \mathbf{E} \left[ \sup_{t \in [0, T]} |X_{t \wedge \theta_J}^{(j)} - \bar{X}_{t \wedge \theta_J}^{(j)}|^p \right] + \left( \mathbf{E} \left[ \sup_{t \in [0, T]} |X_t^{(j)} - \bar{X}_t^{(j)}|^q \right] \right)^{\frac{p}{q}} \left( \mathbf{P}[\theta_J(R) \leq T] \right)^{\frac{q-p}{q}}. \end{aligned} \quad (5.3)$$

**Bounding the first term in (5.3).** We again employ an adapted version of Sznitman's argument. We have

$$\begin{aligned} \left| X_{t \wedge \theta_J}^{(j)} - \bar{X}_{t \wedge \theta_J}^{(j)} \right|^p &\leq \left| \int_0^{t \wedge \theta_J} b \left( X_s^{(j)}, \mu_s^J \right) - b \left( \bar{X}_s^{(j)}, \bar{\rho}_s \right) ds \right|^p \\ &\quad + \left| \int_0^{t \wedge \theta_J} \sigma \left( X_s^{(j)}, \mu_s^J \right) - \sigma \left( \bar{X}_s^{(j)}, \bar{\rho}_s \right) dW_s \right|^p. \end{aligned}$$

Let us introduce the martingale

$$M_t = \int_0^t \sigma \left( X_s^{(j)}, \mu_s^J \right) - \sigma \left( \bar{X}_s^{(j)}, \bar{\rho}_s \right) dW_s.$$

By Doob's optional stopping theorem [34, Theorem 3.3], see also [21, Equation 2.29, p.285], the stopped process  $(M_{t \wedge \theta_J})_{t \geq 0}$  is a martingale, with a quadratic variation process given by the stopped quadratic variation process of  $M_t$ , i.e. by  $(\langle M \rangle_{t \wedge \theta_J})_{t \geq 0}$ . Therefore, by the Burkholder–Davis–Gundy inequality, we have for all  $t \in [0, T]$  that

$$\begin{aligned} \mathbf{E} \left[ \sup_{s \in [0, t]} \left| X_{s \wedge \theta_J}^{(j)} - \bar{X}_{s \wedge \theta_J}^{(j)} \right|^p \right] &\leq (2T)^{p-1} \mathbf{E} \int_0^{t \wedge \theta_J} \left| b \left( X_s^{(j)}, \mu_s^J \right) - b \left( \bar{X}_s^{(j)}, \bar{\rho}_s \right) \right|^p ds \\ &\quad + C_{\text{BDG}} 2^{p-1} T^{\frac{p}{2}-1} \mathbf{E} \int_0^{t \wedge \theta_J} \left\| \sigma \left( X_s^{(j)}, \mu_s^J \right) - \sigma \left( \bar{X}_s^{(j)}, \bar{\rho}_s \right) \right\|_{\mathbb{F}}^p ds. \end{aligned} \quad (5.4)$$

From the triangle inequality we obtain that

$$\begin{aligned} \frac{1}{2^{p-1}} \mathbf{E} \int_0^{t \wedge \theta_J} \left| b \left( X_s^{(j)}, \mu_s^J \right) - b \left( \bar{X}_s^{(j)}, \bar{\rho}_s \right) \right|^p ds &\leq \int_0^t \mathbf{E} \left| b \left( X_{s \wedge \theta_J}^{(j)}, \mu_{s \wedge \theta_J}^J \right) - b \left( \bar{X}_{s \wedge \theta_J}^{(j)}, \bar{\rho}_{s \wedge \theta_J} \right) \right|^p ds \\ &\quad + \int_0^t \mathbf{E} \left| b \left( \bar{X}_s^{(j)}, \bar{\rho}_s \right) - b \left( \bar{X}_s^{(j)}, \bar{\rho}_s \right) \right|^p ds. \end{aligned} \quad (5.5)$$

Similarly, for the diffusion term, we have

$$\begin{aligned} \frac{1}{2^{p-1}} \mathbf{E} \int_0^{t \wedge \theta_J} \left\| \sigma \left( X_s^{(j)}, \mu_s^J \right) - \sigma \left( \bar{X}_s^{(j)}, \bar{\rho}_s \right) \right\|_{\mathbb{F}}^p ds &\leq \int_0^t \mathbf{E} \left\| \sigma \left( X_{s \wedge \theta_J}^{(j)}, \mu_{s \wedge \theta_J}^J \right) - \sigma \left( \bar{X}_{s \wedge \theta_J}^{(j)}, \bar{\rho}_{s \wedge \theta_J} \right) \right\|_{\mathbb{F}}^p ds \\ &\quad + \int_0^t \mathbf{E} \left\| \sigma \left( \bar{X}_s^{(j)}, \bar{\rho}_s \right) - \sigma \left( \bar{X}_s^{(j)}, \bar{\rho}_s \right) \right\|_{\mathbb{F}}^p ds. \end{aligned} \quad (5.6)$$

Next, we need to bound the terms on the right-hand side of (5.5) and (5.6). Since the drift terms and the diffusion terms are bounded in a similar manner, we only present the bounds for the diffusion terms.

- For the first diffusion term, we have that

$$\left\| \sigma \left( X_{s \wedge \theta_J}^{(j)}, \mu_{s \wedge \theta_J}^J \right) - \sigma \left( \bar{X}_{s \wedge \theta_J}^{(j)}, \bar{\rho}_{s \wedge \theta_J} \right) \right\|_{\mathbb{F}}^p = (2\lambda^{-1})^{\frac{p}{2}} \left\| \sqrt{C_\beta \left( \mu_{s \wedge \theta_J}^J \right)} - \sqrt{C_\beta \left( \bar{\rho}_{s \wedge \theta_J} \right)} \right\|_{\mathbb{F}}^p.$$

Using Corollary 3.3, together with the fact

$$\mathbf{E} \left[ W_p \left( \mu_{s \wedge \theta_J}^J, \bar{\rho}_{s \wedge \theta_J} \right)^p \right] \leq \mathbf{E} \left[ \frac{1}{J} \sum_{j=1}^J \left| X_{s \wedge \theta_J}^{(j)} - \bar{X}_{s \wedge \theta_J}^{(j)} \right|^p \right] = \mathbf{E} \left| X_{s \wedge \theta_J}^{(j)} - \bar{X}_{s \wedge \theta_J}^{(j)} \right|^p,$$

we deduce that

$$\mathbf{E} \left\| \sigma \left( X_{s \wedge \theta_J}^{(j)}, \mu_{s \wedge \theta_J}^J \right) - \sigma \left( \bar{X}_{s \wedge \theta_J}^{(j)}, \bar{\rho}_{s \wedge \theta_J} \right) \right\|_{\mathbb{F}}^p \leq C \mathbf{E} \left| X_{s \wedge \theta_J}^{(j)} - \bar{X}_{s \wedge \theta_J}^{(j)} \right|^p.$$

- For the other diffusion term, we use [Lemma 3.8](#). The assumptions of this lemma are satisfied since, on the one hand,  $\bar{\rho}_s \in \mathcal{P}_q(\mathbf{R}^d)$  by [Theorem 2.4](#) and  $2p < q$ , and on the other hand, it holds that  $\mathcal{C}_\beta(\bar{\rho}_s) \succ 0$  by virtue of [Proposition 3.9](#). We deduce that

$$\mathbf{E} \left\| \sigma \left( \bar{X}_s^{(j)}, \bar{\mu}_s^J \right) - \sigma \left( \bar{X}_s^{(j)}, \bar{\rho}_s \right) \right\|_{\mathbb{F}}^p = (2\lambda^{-1})^{\frac{p}{2}} \mathbf{E} \left\| \sqrt{\mathcal{C}_\beta(\bar{\mu}_s^J)} - \sqrt{\mathcal{C}_\beta(\bar{\rho}_s)} \right\|_{\mathbb{F}}^p \leq C J^{-\frac{p}{2}}.$$

Combining the bounds on both terms, we deduce that

$$\mathbf{E} \int_0^{t \wedge \theta_J} \left\| \sigma \left( X_s^{(j)}, \mu_s^J \right) - \sigma \left( \bar{X}_s^{(j)}, \bar{\rho}_s \right) \right\|_{\mathbb{F}}^p ds \leq C J^{-\frac{p}{2}} + C \int_0^t \mathbf{E} \left[ \sup_{u \in [0, s]} \left| X_{u \wedge \theta_J}^{(j)} - \bar{X}_{u \wedge \theta_J}^{(j)} \right|^p \right] ds.$$

Applying a similar reasoning to the drift terms, we finally obtain from (5.4) that

$$\mathbf{E} \left[ \sup_{s \in [0, t]} \left| X_{s \wedge \theta_J}^{(j)} - \bar{X}_{s \wedge \theta_J}^{(j)} \right|^p \right] \leq C J^{-\frac{p}{2}} + C \int_0^t \mathbf{E} \left[ \sup_{u \in [0, s]} \left| X_{u \wedge \theta_J}^{(j)} - \bar{X}_{u \wedge \theta_J}^{(j)} \right|^p \right] ds.$$

By Grönwall's inequality, this implies that

$$\mathbf{E} \left[ \sup_{t \in [0, T]} \left| X_{t \wedge \theta_J}^{(j)} - \bar{X}_{t \wedge \theta_J}^{(j)} \right|^p \right] \leq C J^{-\frac{p}{2}}.$$

**Bounding the second term in (5.3).** We have by [Lemma 3.5](#) and (2.6) that

$$\mathbf{E} \left[ \sup_{t \in [0, T]} \left| X_t^{(j)} - \bar{X}_t^{(j)} \right|^q \right] \leq 2^{q-1} \left( \mathbf{E} \left[ \sup_{t \in [0, T]} \left| X_t^{(j)} \right|^q \right] + \mathbf{E} \left[ \sup_{t \in [0, T]} \left| \bar{X}_t^{(j)} \right|^q \right] \right) \leq 2^q \kappa.$$

In addition, by [Lemma 2.5](#) it holds that

$$\mathbf{P}[\theta_J(R) \leq T] = \mathbf{P} \left[ \sup_{t \in [0, T]} \frac{1}{J} \sum_{j=1}^J \left| \bar{X}_t^{(j)} \right|^p \geq R \right] \leq C J^{-2}.$$

**Conclusion.** We finally obtain from (5.3) that

$$\mathbf{E} \left[ \sup_{t \in [0, T]} \left| X_t^{(j)} - \bar{X}_t^{(j)} \right|^p \right] \leq C \left( J^{-\frac{p}{2}} + J^{-\frac{2(q-p)}{q}} \right),$$

which concludes the proof.  $\square$

## 5.2 Handling the case where $u > \ell > 0$ in [Assumption H](#)

Our main results stand only when  $f \in \mathcal{A}(s, \ell, u)$  with  $u = \ell$ . This is a convenient setting because it then holds by [Proposition A.3](#) that the  $p$ -th moment of the reweighted measure  $L_\beta \mu$  is controlled by the  $p$ -th moment of the  $\mu$ , which is key in order to prove well-posedness of the mean-field system, to obtain moment estimates for empirical measures, and, most importantly, to improve the Wasserstein stability estimates for the weighted mean and weighted covariance, see [Corollary 3.3](#). In the interest of simplicity, we refrained from stating precise results for the cases  $u > \ell = 0$  and  $u > \ell > 0$ . The former remains open, but we note that a quantitative mean-field estimate can still be obtained in the latter setting. The main ingredients to this end are the following:

- Moment estimates for the empirical measure and well-posedness of the mean-field equation can be

obtained by using [Proposition A.3](#) together with the Bihari–LaSalle inequality instead of Grönwall’s inequality, using the same reasoning as in [Lemma 3.5](#) and [Section 4](#), respectively. In contrast with the setting  $\ell = u$ , these results hold only up to some finite time  $T < T_+$ , where  $T_+ \in (0, \infty)$  is the time at which the bounds through the Bihari–LaSalle inequality become singular.

- In order to obtain a mean-field result up to some time  $T < T_+$ , the approach presented in [Subsection 5.1](#) can be used, this time with the stopping time  $\theta_J := \tau_J \wedge \bar{\tau}_J$ , where

$$\begin{aligned}\tau_J(R) &= \inf \left\{ t \geq 0 : \mu_t^J[|x|^p] \geq R \right\}, \\ \bar{\tau}_J(R) &= \inf \left\{ t \geq 0 : \bar{\mu}_t^J[|x|^p] \geq R \right\}.\end{aligned}$$

Employing this stopping time instead of (5.2) allows for using the basic stability estimates for the weighted moments of [Proposition 3.1](#), which are valid even when  $\ell < u$ , instead of the improved stability estimates in [Corollary 3.3](#), which require that  $\ell = u$ . In order to complete the proof as in [Subsection 5.1](#), it remains to bound the probability  $\mathbf{P}[\theta_J \leq T]$ , which can be achieved by noticing that

$$\mathbf{P}[\theta_J(R) \leq T] = \mathbf{P}[\tau_J(R) \leq T < \bar{\tau}_J(R)] + \mathbf{P}[\bar{\tau}_J(R) \leq T].$$

Using the almost sure continuity of the solution to the interacting particle system by [Theorem 2.2](#), together with the triangle inequality, we bound the first probability as follows:

$$\begin{aligned}\mathbf{P}[\tau_J \leq T < \bar{\tau}_J] &\leq \mathbf{P} \left[ \sup_{t \in [0, T]} \frac{1}{J} \sum_{j=1}^J |X_{t \wedge \theta_J}^{(j)}|^p = R \right] \\ &\leq \mathbf{P} \left[ \frac{2^{p-1}}{J} \sum_{j=1}^J \sup_{t \in [0, T]} |\bar{X}_t^{(j)}|^p + \frac{2^{p-1}}{J} \sum_{j=1}^J \sup_{t \in [0, T]} |X_{t \wedge \theta_J}^{(j)} - \bar{X}_{t \wedge \theta_J}^{(j)}|^p \geq R \right] \\ &\leq \mathbf{P} \left[ \frac{1}{J} \sum_{j=1}^J \sup_{t \in [0, T]} |\bar{X}_t^{(j)}|^p \geq \frac{R}{2^p} \right] + \mathbf{P} \left[ \frac{1}{J} \sum_{j=1}^J \sup_{t \in [0, T]} |X_{t \wedge \theta_J}^{(j)} - \bar{X}_{t \wedge \theta_J}^{(j)}|^p \geq \frac{R}{2^p} \right],\end{aligned}$$

where we used that  $\mathbf{P}[A + B \geq k] \leq \mathbf{P}[A \geq k/2] + \mathbf{P}[B \geq k/2]$  for any two real-valued random variables  $A$  and  $B$ , because  $\{A + B \geq k\} \subset \{A \geq k/2\} \cup \{B \geq k/2\}$ . A similar trick, of bounding the probability of an event associated with a stochastic system in terms of a similar event associated with a simpler system and the discrepancy between the two systems, is employed in [\[20, Theorem 2.10\]](#). The probability  $\mathbf{P}[\theta_J \leq T]$  can then be bounded as follows:

$$\begin{aligned}\mathbf{P}[\theta_J(R) \leq T] &= \mathbf{P}[\tau_J(R) \leq T < \bar{\tau}_J] + \mathbf{P}[\bar{\tau}_J(R) \leq T] \\ &\leq 2\mathbf{P} \left[ \frac{1}{J} \sum_{j=1}^J \sup_{t \in [0, T]} |\bar{X}_t^{(j)}|^p \geq \frac{R}{2^p} \right] + \mathbf{P} \left[ \frac{1}{J} \sum_{j=1}^J \sup_{t \in [0, T]} |X_{t \wedge \theta_J}^{(j)} - \bar{X}_{t \wedge \theta_J}^{(j)}|^p \geq \frac{R}{2^p} \right].\end{aligned}$$

Using Markov’s inequality for the second term, we then obtain

$$\mathbf{P}[\theta_J(R) \leq T] \leq 2\mathbf{P} \left[ \frac{1}{J} \sum_{j=1}^J \sup_{t \in [0, T]} |\bar{X}_t^{(j)}|^p \geq \frac{R}{2^p} \right] + \frac{2^p}{R} \mathbf{E} \left[ \frac{1}{J} \sum_{j=1}^J \sup_{t \in [0, T]} |X_{t \wedge \theta_J}^{(j)} - \bar{X}_{t \wedge \theta_J}^{(j)}|^p \right]. \quad (5.7)$$

The first term is bounded from above by  $CJ^{-2}$  by [Lemma 2.5](#), whereas the second term can be bounded by  $CJ^{-\frac{p}{2}}$  as in the proof in [Subsection 5.1](#). Thus, the probability  $\mathbf{P}[\theta_J(R) \leq T]$  scales as  $J^{-\min\{2, \frac{p}{2}\}}$ ,



instead of  $J^{-2}$  in the case where  $u = \ell$ . As a consequence, we obtain a mean-field result of the type

$$\forall J \in \mathbf{N}^+, \quad \mathbf{E} \left[ \sup_{t \in [0, T]} \left| X_t^{(j)} - \bar{X}_t^{(j)} \right|^p \right] \leqslant C J^{-\min\{2, \frac{p}{2}\} \frac{q-p}{q}}.$$

## A Further technical results

### A.1 Stability estimates for ratios of integrals

**Lemma A.1.** *For a real finite-dimensional vector space  $\mathcal{V}$  with norm  $\|\bullet\|$ , let  $g: \mathbf{R}^d \rightarrow \mathcal{V}$  and  $h: \mathbf{R}^d \rightarrow (0, \infty)$  be functions such that the following condition is satisfied for some  $\varsigma \geqslant 0$  and  $L > 0$ :*

$$\forall x, y \in \mathbf{R}^d, \quad \|g(x) - g(y)\| \vee |h(x) - h(y)| \leqslant L(1 + |x| + |y|)^\varsigma |x - y|. \quad (\text{A.1})$$

Then for all  $p \geqslant \varsigma + 1$  and all  $R > 0$ , there exist a constant  $C = C(p, R)$  such that

$$\forall (\mu, \nu) \in \mathcal{P}_{p, R}(\mathbf{R}^d) \times \mathcal{P}_{p, R}(\mathbf{R}^d), \quad \left\| \frac{\mu[g]}{\mu[h]} - \frac{\nu[g]}{\nu[h]} \right\| \leqslant C W_p(\mu, \nu).$$

*Proof.* Fix  $p \geqslant \varsigma + 1$  and  $R > 0$ .

**Step 1.** Let us first generalize [11, Lemma 3.1] by showing that there exists a constant  $\eta = \eta(p, R) > 0$  such that

$$\forall \mu \in \mathcal{P}_{p, R}(\mathbf{R}^d), \quad \int_{\mathbf{R}^d} h \, d\mu \geqslant \eta. \quad (\text{A.2})$$

Define  $B_K := \{x \in \mathbf{R}^d : |x| \leqslant K\}$  where  $K := 2^{1/p} R$ . By Markov's inequality, we have

$$\forall \mu \in \mathcal{P}_{p, R}(\mathbf{R}^d), \quad \mu(B_K) \geqslant 1 - \mu(\mathbf{R}^d \setminus B_K) \geqslant 1 - \frac{R^p}{K^p} = \frac{1}{2}.$$

Therefore we obtain

$$\forall \mu \in \mathcal{P}_{p, R}(\mathbf{R}^d), \quad \int_{\mathbf{R}^d} h \, d\mu \geqslant \int_{B_K} h \, d\mu \geqslant \frac{1}{2} \min\{h(x) : x \in B_K\} =: \eta > 0,$$

since  $h$  is continuous by (A.1). This proves (A.2).

**Step 2.** Fix probability measures  $\mu, \nu \in \mathcal{P}_{p, R}$ . We omit the integration domains for conciseness. For any coupling  $\pi \in \Pi(\mu, \nu)$ , we have by (A.1) and (A.2) that

$$\begin{aligned} \left\| \frac{\mu[g]}{\mu[h]} - \frac{\nu[g]}{\nu[h]} \right\| &\leqslant \left\| \frac{\iint g(x) - g(y) \pi(dx \, dy)}{\mu[h]} \right\| + \left\| \int g \, d\nu \right\| \cdot \left| \frac{\iint h(x) - h(y) \pi(dx \, dy)}{\mu[h]\nu[h]} \right| \\ &\leqslant L \left( \eta^{-1} + \eta^{-2} \int \|g\| \, d\nu \right) \iint (1 + |x| + |y|)^\varsigma |x - y| \pi(dx \, dy). \end{aligned}$$

By (A.1), there exists  $C > 0$  such that  $\|g(x)\| \leqslant C(1 + |x|)^{\varsigma+1}$  for all  $x \in \mathbf{R}^d$ . The claim then follows by first applying Hölder's inequality to obtain the bound

$$\iint (1 + |x| + |y|)^\varsigma |x - y| \pi(dx \, dy) \leqslant \left( \iint (1 + |x| + |y|)^{\frac{sp}{p-1}} \pi(dx \, dy) \right)^{\frac{p-1}{p}} \left( \iint |x - y|^p \pi(dx \, dy) \right)^{\frac{1}{p}},$$

noting that  $\frac{sp}{p-1} \leqslant p$  under the assumption that  $p \geqslant \varsigma + 1$ , and then infimizing over  $\pi \in \Pi(\mu, \nu)$ .  $\square$

**Lemma A.2.** Let  $g: \mathbf{R}^d \rightarrow \mathbf{R}^{m \times n}$ ,  $h: \mathbf{R}^d \rightarrow (0, \infty)$  and  $\mathcal{S}: \mathcal{P}(\mathbf{R}^d) \rightarrow \mathbf{R}^n$  be functions such that the following conditions hold.

(a) There is  $\varsigma \geq 0$  and  $L > 0$  such that

$$\forall x, y \in \mathbf{R}^d, \quad \|g(x) - g(y)\| \vee |h(x) - h(y)| \leq L(1 + |x| + |y|)^\varsigma |x - y|.$$

(b) The map  $\mathcal{S}: \mathcal{P}(\mathbf{R}^d) \rightarrow \mathbf{R}^n$  satisfies

$$\forall (\mu, \nu) \in \mathcal{P}_{p,R}(\mathbf{R}^d) \times \mathcal{P}_{p,R}(\mathbf{R}^d), \quad |\mathcal{S}(\mu) - \mathcal{S}(\nu)| \leq L_{\mathcal{S}} W_p(\mu, \nu).$$

Then, for all  $R > 0$  and  $p \geq 2\varsigma + 2$ , there exist a constant  $C > 0$  such that for all  $\mu, \nu \in \mathcal{P}_{p,R}$  and all couplings  $\pi \in \Pi(\mu, \nu)$ , it holds that

$$\left( \iint_{\mathbf{R}^d \times \mathbf{R}^d} \left\| \frac{g(x)\mathcal{S}(\mu)}{\sqrt{\mu[h]}} - \frac{g(y)\mathcal{S}(\nu)}{\sqrt{\nu[h]}} \right\|^2 \pi(dx dy) \right)^{1/2} \leq C \left( \int |x - y|^p \pi(dx dy) \right)^{1/p}. \quad (\text{A.3})$$

*Proof.* Fix  $p \geq 2\varsigma + 2$  and  $R > 0$ . Again, there is a constant  $\eta = \eta(R, p) > 0$  such that (A.2) holds. From the triangle inequality we obtain

$$\begin{aligned} \left\| \frac{g(x)\mathcal{S}(\mu)}{\sqrt{\mu[h]}} - \frac{g(y)\mathcal{S}(\nu)}{\sqrt{\nu[h]}} \right\|_{L^2(\pi)} &\leq \left\| \frac{g(x)}{\sqrt{\mu[h]}} (\mathcal{S}(\mu) - \mathcal{S}(\nu)) \right\|_{L^2(\pi)} + \left\| (g(x) - g(y)) \frac{\mathcal{S}(\nu)}{\sqrt{\mu[h]}} \right\|_{L^2(\pi)} \\ &\quad + \left\| g(y)\mathcal{S}(\nu) \left( \frac{1}{\sqrt{\mu[h]}} - \frac{1}{\sqrt{\nu[h]}} \right) \right\|_{L^2(\pi)}. \end{aligned}$$

where, by a slight abuse of notation, we wrote  $g(x)$  and  $g(y)$  in order to refer to the functions  $(x, y) \mapsto g(x)$  and  $(x, y) \mapsto g(y)$ , respectively. In the rest of the proof, we denote by  $C$  any constant depending only on the parameters  $(R, L, \varsigma, \eta, p)$ .

**Bounding the first term.** It follows from (a) that  $\|g(x)\|^2 \leq C(1 + |x|)^{2\varsigma+2}$  for all  $x \in \mathbf{R}^d$ , for a constant  $C$ . Using (b), we obtain the bound

$$\left( \iint \left\| \frac{g(x)}{\sqrt{\mu[h]}} (\mathcal{S}(\mu) - \mathcal{S}(\nu)) \right\|^2 \pi(dx dy) \right)^{1/2} \leq \sqrt{\frac{\mu[\|g(x)\|^2]}{\mu[h]}} |\mathcal{S}(\mu) - \mathcal{S}(\nu)| \leq \sqrt{\frac{C}{\eta}} L_{\mathcal{S}} W_p(\mu, \nu).$$

**Bounding the second term.** For the second term, we use (a) to find

$$\forall x, y \in \mathbf{R}^d, \quad |g(x) - g(y)|^2 \leq L^2 (1 + |x| + |y|)^{2\varsigma} |x - y|^2.$$

Using Hölder's inequality, we therefore obtain

$$\begin{aligned} \|g(x) - g(y)\|_{L^2(\pi)} &\leq L \left( \int (1 + |x| + |y|)^{\frac{2\varsigma p}{p-2}} \pi(dx dy) \right)^{\frac{p-2}{2p}} \left( \int |x - y|^p \pi(dx dy) \right)^{\frac{1}{p}} \\ &\leq C \left( \int |x - y|^p \pi(dx dy) \right)^{\frac{1}{p}}, \end{aligned}$$

where we used that  $\frac{2sp}{p-2} \leq p$  since  $p \geq 2 + 2\varsigma$ . In addition, from (b) we have

$$\forall \mu \in \mathcal{P}_{p,R}(\mathbf{R}^d), \quad |\mathcal{S}(\mu)| \leq |\mathcal{S}(\delta_0)| + L_S W_p(\delta_0, \mu) \leq |\mathcal{S}(\delta_0)| + L_S R. \quad (\text{A.4})$$

and so we conclude that

$$\left\| \left( g(x) - g(y) \right) \frac{\mathcal{S}(\nu)}{\sqrt{\mu[h]}} \right\|_{L^2(\pi)} \leq C \left( \int |x - y|^p \pi(\mathrm{d}x \mathrm{d}y) \right)^{1/p}. \quad (\text{A.5})$$

**Bounding the third term.** Using again (A.4), we have

$$\left\| g(y) \mathcal{S}(\nu) \left( \frac{1}{\sqrt{\mu[h]}} - \frac{1}{\sqrt{\nu[h]}} \right) \right\|_{L^2(\pi)} = \|g\|_{L^2(\mu)} |\mathcal{S}(\nu)| \left| \frac{1}{\sqrt{\mu[h]}} - \frac{1}{\sqrt{\nu[h]}} \right| \leq C \left| \frac{\sqrt{\mu[h]} - \sqrt{\nu[h]}}{\sqrt{\mu[h]}\sqrt{\nu[h]}} \right|.$$

Since  $\sqrt{a} - \sqrt{b} = (a - b)/(\sqrt{a} + \sqrt{b})$  for all  $a, b > 0$ , we deduce that

$$\begin{aligned} \left\| g(x) \mathcal{S}(\nu) \left( \frac{1}{\sqrt{\mu[h]}} - \frac{1}{\sqrt{\nu[h]}} \right) \right\|_{L^2(\pi)} &\leq C \eta^{-\frac{3}{2}} \iint |h(x) - h(y)| \pi(\mathrm{d}x \mathrm{d}y) \\ &\leq C \iint (1 + |x| + |y|)^\varsigma |x - y| \pi(\mathrm{d}x \mathrm{d}y). \end{aligned}$$

By employing Hölder's inequality, the integral can be bounded by the right-hand side of (A.5) up to multiplication by a constant.

**Conclusion.** The claim follows then by combining the bounds on the three terms, and noting, for the first term, that  $W_p(\mu, \nu)$  is bounded from above by the bracket on the right-hand side of (A.3) by definition of the Wasserstein distance.  $\square$

## A.2 Bound on weighted moments

If the function  $f$  is bounded from below and above, it is trivial to show that

$$\forall \mu \in \mathcal{P}_p(\mathbf{R}^d), \quad \frac{\int_{\mathbf{R}^d} |x|^p e^{-\beta f(x)} \mu(\mathrm{d}x)}{\int_{\mathbf{R}^d} e^{-\beta f(x)} \mu(\mathrm{d}x)} \leq C \int_{\mathbf{R}^d} |x|^p \mu(\mathrm{d}x).$$

for a constant  $C$  depending only on  $\beta$  and  $f$ . When  $f$  is unbounded, proving an inequality of this type is more involved. The following is a generalization of [11, Lemma 3.3].

**Proposition A.3.** *Suppose that  $f$  is bounded from below by  $f_\star$  and that there are real numbers  $u \geq \ell > 0$  and positive constants  $c_u, c_\ell$  such that*

$$\forall x \in \mathbf{R}^d, \quad \begin{cases} f(x) - f_\star \leq c_u (|x|^u + 1) \\ f(x) - f_\star \geq c_\ell (|x|^\ell - 1). \end{cases} \quad (\text{A.6})$$

Then for all  $p, q > 0$  there are constants  $C_1$  and  $C_2$  depending on  $(p, q, u, \ell, c_u, c_\ell)$  such that

$$\forall \mu \in \mathcal{P}_q(\mathbf{R}^d), \quad \frac{\int_{\mathbf{R}^d} |x|^p e^{-\beta f(x)} \mu(\mathrm{d}x)}{\int_{\mathbf{R}^d} e^{-\beta f(x)} \mu(\mathrm{d}x)} \leq \left( C_1 + C_2 \int_{\mathbf{R}^d} |x|^q \mu(\mathrm{d}x) \right)^{\frac{pu}{q\ell}} \quad (\text{A.7})$$

Furthermore, if  $u = \ell$  and  $p \leq q$ , then this statement holds with  $C_1 = 0$ .

The proof relies on the following lemma.

**Lemma A.4.** For all  $p, q, \beta > 0$ , there is  $C = C(p, q, \beta)$  such that

$$\forall \nu \in \mathcal{P}_q(\mathbf{R}^+), \quad \frac{\int_{\mathbf{R}^+} y^p e^{-\beta y} \nu(dy)}{\int_{\mathbf{R}^+} e^{-\beta y} \nu(dy)} \leq C \left( 1 + \int_{\mathbf{R}^+} y^q \nu(dy) \right)^{\frac{p}{q}}, \quad (\text{A.8})$$

where  $\mathbf{R}^+ := [0, \infty)$ . The constant is given by

$$C(p, q, \beta) = \left( 1 - \frac{1}{2^q} \right)^{-1} (p\beta^{-1})^p e^{2\beta-p}.$$

*Proof.* Fix  $\nu \in \mathcal{P}(\mathbf{R}^+)$ . If the right-hand side of (A.8) is infinite, then the inequality (A.8) holds with any constant  $C$ , so we assume from now on that

$$\lambda := \left( \int_{\mathbf{R}^+} y^q \nu(dy) \right)^{\frac{1}{q}} < \infty.$$

**Step 1.** Suppose first that  $\lambda \leq 1$ . Then, by Markov's inequality, it holds that

$$\nu([0, 2]) = 1 - \int_{\mathbf{R}^+ \setminus [0, 2]} \nu(dy) \geq 1 - \frac{1}{2^q} \int_{\mathbf{R}^+} y^q \nu(dy) \geq 1 - \frac{1}{2^q}.$$

Using this inequality and the bound  $y^p e^{-\beta y} \leq (p\beta^{-1})^p e^{-p}$ , which is valid for all  $y \in \mathbf{R}^+$  and obtained by maximization of  $y \mapsto y^p e^{-\beta y}$ , we have that

$$\frac{\int_{\mathbf{R}^+} y^p e^{-\beta y} \nu(dy)}{\int_{\mathbf{R}^+} e^{-\beta y} \nu(dy)} \leq \frac{\int_{\mathbf{R}^+} (p\beta^{-1})^p e^{-p} \nu(dy)}{\int_{[0, 2]} e^{-\beta y} \nu(dy)} \leq \left( 1 - \frac{1}{2^q} \right)^{-1} (p\beta^{-1})^p e^{2\beta-p}.$$

It follows that (A.8) is satisfied in this case.

**Step 2.** Assume now that  $\lambda \geq 1$ , and let  $\pi$  denote the pushforward of  $\nu$  under the contraction  $x \mapsto \lambda^{-1}x$ . Then, applying a change of variable, using **Step 3** and using **Step 1** step, we have that

$$\begin{aligned} \frac{\int_{\mathbf{R}^+} y^p e^{-\beta y} \nu(dy)}{\int_{\mathbf{R}^+} e^{-\beta y} \nu(dy)} &= \lambda^p \frac{\int_{\mathbf{R}^+} z^p e^{-\beta \lambda z} \pi(dz)}{\int_{\mathbf{R}^+} e^{-\beta \lambda z} \pi(dz)} \\ &\leq \lambda^p \frac{\int_{\mathbf{R}^+} z^p e^{-\beta z} \pi(dz)}{\int_{\mathbf{R}^+} e^{-\beta z} \pi(dz)} \leq \lambda^p C(p, q, \beta), \end{aligned} \quad (\text{A.9})$$

which is the required inequality.

**Step 3.** In the first inequality of (A.9), we used that, for any increasing function  $g: \mathbf{R}^+ \rightarrow \mathbf{R}$ , the function

$$F(\gamma) := \frac{\int_{\mathbf{R}^+} g(z) e^{-\gamma z} \pi(dz)}{\int_{\mathbf{R}^+} e^{-\gamma z} \pi(dz)}$$

is a non-increasing function of  $\gamma$ . This follows from [12, Lemma A.3], and could also be proved by showing that  $F'(\gamma)$  is non-positive for all  $\gamma \geq 0$ . A simple calculation indeed gives that

$$F'(\gamma) = - \int_{\mathbf{R}^+} g(z) (z - \bar{z}) L_\gamma \pi(dz), \quad L_\gamma \pi(dz) := \frac{e^{-\gamma z} \pi(dz)}{\int_{\mathbf{R}^+} e^{-\gamma z} \pi(dz)}, \quad \bar{z} := \int_{\mathbf{R}^+} z L_\gamma \pi(dz).$$

By decomposing the integration domain into the union  $[0, \bar{z}) \cup \{\bar{z}\} \cup (\bar{z}, \infty)$ , and using the fact that  $g$  is non-decreasing, we have

$$\begin{aligned} F'(\gamma) &= - \int_{[0, \bar{z})} g(z) (z - \bar{z}) \pi(dz) - \int_{(\bar{z}, \infty)} g(z) (z - \bar{z}) L_\gamma \pi(dz) \\ &\leq -g(\bar{z}) \int_{[0, \bar{z}) \cup (\bar{z}, \infty)} (z - \bar{z}) L_\gamma \pi(dz) = 0, \end{aligned}$$

which completes the proof.  $\square$

*Proof of Proposition A.3.* We assume without loss of generality that  $f_\star = 0$ . By the second inequality in (A.6), we have that

$$|x|^p = \left(|x|^\ell\right)^{\frac{p}{\ell}} \leq \left(\frac{f(x)}{c_\ell} + 1\right)^{\frac{p}{\ell}}. \quad (\text{A.10})$$

For convenience, we denote by  $C$  any constant depending only on  $(p, q, \ell, u, c_\ell, c_u)$ . From (A.10) we deduce that

$$\begin{aligned} \frac{\int_{\mathbf{R}^d} |x|^p e^{-\beta f(x)} \mu(dx)}{\int_{\mathbf{R}^d} e^{-\beta f(x)} \mu(dx)} &\leq C \frac{\int_{\mathbf{R}^d} (f(x) + 1)^{\frac{p}{\ell}} e^{-\beta f(x)} \mu(dx)}{\int_{\mathbf{R}^d} e^{-\beta f(x)} \mu(dx)} \\ &= C \frac{\int_{\mathbf{R}^+} (y + 1)^{\frac{p}{\ell}} e^{-\beta y} f_\# \mu(dy)}{\int_{\mathbf{R}^+} e^{-\beta y} f_\# \mu(dy)}, \end{aligned}$$

where  $f_\# \mu \in \mathcal{P}(\mathbf{R}^+)$  is the pushforward of  $\mu$  by  $f$ . Using Lemma A.4 and then the first inequality in (A.6), we obtain

$$\begin{aligned} \frac{\int_{\mathbf{R}^d} |x|^p e^{-\beta f(x)} \mu(dx)}{\int_{\mathbf{R}^d} e^{-\beta f(x)} \mu(dx)} &\leq C \left(1 + \int_{\mathbf{R}^+} |y|^{\frac{q}{u}} f_\# \mu(dy)\right)^{\frac{pu}{q\ell}} \\ &= C \left(1 + \int_{\mathbf{R}^d} |f(x)|^{\frac{q}{u}} \mu(dx)\right)^{\frac{pu}{q\ell}} \leq C \left(1 + \int_{\mathbf{R}^d} |x|^q \mu(dx)\right)^{\frac{pu}{q\ell}}, \end{aligned}$$

which completes the proof of (A.7).

**Proof that (A.7) holds with  $C_1 = 0$  when  $\ell = u$  and  $q \geq p$ .** The case  $q > p$  follows from the case  $q = p$ , so we assume the latter. Suppose for contradiction that this was not true. Then there would exist a sequence  $\{\mu_i\}_{i \in \mathbf{N}}$  of probability measures such that

$$\frac{\int_{\mathbf{R}^d} |x|^p e^{-\beta f(x)} \mu_i(dx)}{\int_{\mathbf{R}^d} e^{-\beta f(x)} \mu_i(dx)} \geq i \int_{\mathbf{R}^d} |x|^p \mu_i(dx). \quad (\text{A.11})$$

By the first part of the proof, this inequality implies that

$$\frac{C_1 + C_2 \int_{\mathbf{R}^d} |x|^p \mu_i(dx)}{\int_{\mathbf{R}^d} |x|^p \mu_i(dx)} \geq i,$$

and so the integral in this expression necessarily converges to 0 in the limit as  $i \rightarrow \infty$ . In turn, this implies that the sequence  $\{\mu_i\}_{i \in \mathbf{N}}$  converges in distribution to the Dirac measure at 0 when  $i \rightarrow \infty$ . But then it holds that

$$\frac{\int_{\mathbf{R}^d} |x|^p e^{-\beta f(x)} \mu_i(dx)}{\int_{\mathbf{R}^d} e^{-\beta f(x)} \mu_i(dx)} \left( \int_{\mathbf{R}^d} |x|^p \mu_i(dx) \right)^{-1} \leq \frac{e^{-\beta f_\star}}{\int_{\mathbf{R}^d} e^{-\beta f(x)} \mu_i(dx)} \xrightarrow{i \rightarrow \infty} e^{-\beta(f_\star - f(0))},$$

which contradicts (A.11).  $\square$

**Remark A.5.** When  $p = q > 0$ , [Proposition A.3](#) implies that

$$\forall \mu \in \mathcal{P}_p(\mathbf{R}^d), \quad \frac{\int_{\mathbf{R}^d} |x|^p e^{-\beta f(x)} \mu(\mathrm{d}x)}{\int_{\mathbf{R}^d} e^{-\beta f(x)} \mu(\mathrm{d}x)} \leq \left( C_1 + C_2 \int_{\mathbf{R}^d} |x|^p \mu(\mathrm{d}x) \right)^{\frac{u}{\ell}}. \quad (\text{A.12})$$

We motivate in this remark that the exponent  $\frac{u}{\ell}$  is optimal. Indeed, let  $p = 1$  and consider the function  $f: \mathbf{R}^2 \rightarrow \mathbf{R}$  given by  $f(x_1, x_2) = |x_1|^\ell + |x_2|^{u+\varepsilon}$ , for any  $u \geq \ell > 0$  and  $\varepsilon > 0$ . Then for the probability measures

$$\mu_i = \frac{1}{i^u} \delta_{(i^u, 0)} + \left( 1 - \frac{1}{i^u} \right) \delta_{(0, i^\ell)},$$

it holds that

$$\frac{\int_{\mathbf{R}^d} |x| e^{-\beta f(x)} \mu_i(\mathrm{d}x)}{\int_{\mathbf{R}^d} e^{-\beta f(x)} \mu_i(\mathrm{d}x)} \sim i^u \quad \text{and} \quad \int_{\mathbf{R}^2} |x| \mu_i(\mathrm{d}x) \sim i^\ell \quad \text{as } i \rightarrow \infty,$$

where  $\sim$  signifies asymptotic equivalence. This example suggests that [Lemma A.4](#) is sharp.

### A.3 Generalized Wasserstein stability estimates

By adapting the proof of [Proposition 3.1](#), it is simple to prove Wasserstein stability estimates with different values of  $p$  on the left and right sides of the inequality. More precisely, we have the following statement.

**Lemma A.6.** Suppose that  $f \in \mathcal{A}(s, 0, u)$ . Then

- for all  $R > 0$  and  $p > 1$ , there is  $C = C(p, R)$

$$\forall (\mu, \nu) \in \mathcal{P}_{\varrho, R}(\mathbf{R}^d) \times \mathcal{P}_{\varrho, R}(\mathbf{R}^d), \quad |\mathcal{M}_\beta(\mu) - \mathcal{M}_\beta(\nu)| \leq C W_p(\mu, \nu), \quad \varrho := \frac{(s+1)p}{p-1} \quad (\text{A.13})$$

- for all  $R > 0$  and  $p > 1$ , there is  $C = C(p, R)$

$$\forall (\mu, \nu) \in \mathcal{P}_{\varrho, R}(\mathbf{R}^d) \times \mathcal{P}_{\varrho, R}(\mathbf{R}^d), \quad |\mathcal{C}_\beta(\mu) - \mathcal{C}_\beta(\nu)| \leq C W_p(\mu, \nu), \quad \varrho := \frac{(s+2)p}{p-1} \quad (\text{A.14})$$

**Remark A.7.** For  $p = s + 2$ , the bound [\(A.13\)](#) coincides with [\(3.1\)](#), and for  $p = s + 3$ , the bound [\(A.14\)](#) coincides with [\(3.2\)](#).

**Acknowledgments.** We are grateful to Yunan Yang, Arnaud Guyader and Louis-Pierre Chaintron for useful discussions, and to the Lorentz Center (Leiden, the Netherlands) for hosting the workshop ‘‘Purpose-driven particle systems’’ in March 2023, which enabled discussions about this work. FH is supported by start-up funds at the California Institute of Technology. FH was also supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) via project 390685813 - GZ 2047/1 - HCM.

## References

- [1] S. Agapiou, O. Papaspiliopoulos, D. Sanz-Alonso and A. M. Stuart. Importance sampling: intrinsic dimension and computational cost. *Statist. Sci.*, **32**(3):405–431, 2017.
- [2] K. Althaus, I. Papaioannou and E. Ullmann. Consensus-based rare event estimation. *arXiv preprint*, **2304.09077**, 2023.
- [3] H. Araki and S. Yamagami. An inequality for Hilbert-Schmidt norm. *Comm. Math. Phys.*, **81**(1):89–96, 1981.
- [4] H.-O. Bae, S.-Y. Ha, M. Kang, H. Lim, C. Min and J. Yoo. A constrained consensus based optimization algorithm and its application to finance. *Appl. Math. Comput.*, **416**:Paper No. 126726, 10, 2022.

- [5] R. Bailo, A. Barbaro, S. N. Gomes, K. Riedl, T. Roith, C. Totzeck and U. Vaes. CBX: Python and Julia Packages for Consensus-Based Interacting Particle Methods. *Journal of Open Source Software*, **9**(98):6611, 2024.
- [6] R. Bhatia. *Matrix analysis*, volume **169** of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1997.
- [7] R. Bhatia. Matrix factorizations and their perturbations. In volume **197/198**, pages 245–276. 1994. Second Conference of the International Linear Algebra Society (Lisbon, 1992).
- [8] G. Borghi, M. Herty and L. Pareschi. An Adaptive Consensus Based Method for Multi-objective Optimization with Uniform Pareto Front Approximation. *Appl. Math. Optim.*, **88**(2):Paper No. 58, 2023.
- [9] G. Borghi, M. Herty and L. Pareschi. Constrained consensus-based optimization. *SIAM J. Optim.*, **33**(1):211–236, 2023.
- [10] L. Bungert, T. Roith and P. Wacker. Polarized consensus-based dynamics for optimization and sampling. *arXiv preprint*, **2211.05238**, 2022.
- [11] J. A. Carrillo, Y.-P. Choi, C. Totzeck and O. Tse. An analytical framework for consensus-based global optimization method. *Math. Models Methods Appl. Sci.*, **28**(6):1037–1066, 2018.
- [12] J. A. Carrillo, F. Hoffmann, A. M. Stuart and U. Vaes. Consensus-based sampling. *Stud. Appl. Math.*, **148**(3):1069–1140, 2022.
- [13] J. A. Carrillo, C. Totzeck and U. Vaes. Consensus-based optimization and ensemble Kalman inversion for global optimization problems with constraints. *arXiv preprint*, **2111.02970**, 2021.
- [14] J. A. Carrillo, S. Jin, L. Li and Y. Zhu. A consensus-based global optimization method for high dimensional machine learning problems. *ESAIM Control Optim. Calc. Var.*, **27**:Paper No. S5, 22, 2021.
- [15] L.-P. Chaintron and A. Diez. Propagation of chaos: a review of models, methods and applications. I. Models and methods. *Kinet. Relat. Models*, **15**(6):895–1015, 2022.
- [16] L.-P. Chaintron and A. Diez. Propagation of chaos: a review of models, methods and applications. II. Applications. *Kinet. Relat. Models*, **15**(6):1017–1173, 2022.
- [17] Y. S. Chow and H. Teicher. *Probability theory*. Springer Texts in Statistics. Springer-Verlag, New York, third edition, 1997. Independence, interchangeability, martingales.
- [18] C. Cipriani, H. Huang and J. Qiu. Zero-inertia limit: from particle swarm optimization to consensus-based optimization. *SIAM J. Math. Anal.*, **54**(3):3091–3121, 2022.
- [19] P. Doukhan and G. Lang. Evaluation for moments of a ratio with application to regression estimation. *Bernoulli*, **15**(4):1259–1286, 2009.
- [20] M. H. Duong and H. D. Nguyen. Asymptotic analysis for the generalized Langevin equation with singular potentials. *arXiv preprint*, **2305.03637**, 2023.
- [21] S. N. Ethier and T. G. Kurtz. *Markov processes*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons, Inc., New York, 1986. Characterization and convergence.
- [22] M. Fornasier, H. Huang, L. Pareschi and P. Sünnen. Consensus-based optimization on hypersurfaces: well-posedness and mean-field limit. *Math. Models Methods Appl. Sci.*, **30**(14):2725–2751, 2020.
- [23] M. Fornasier, H. Huang, L. Pareschi and P. Sünnen. Consensus-based optimization on the sphere: convergence to global minimizers and machine learning. *J. Mach. Learn. Res.*, **22**:Paper No. 237, 55, 2021.
- [24] M. Fornasier, T. Klock and K. Riedl. Consensus-based optimization methods converge globally. *arXiv preprint*, **2103.15130**, 2021.
- [25] M. Fornasier, T. Klock and K. Riedl. Convergence of anisotropic consensus-based optimization in mean-field law. In J. L. Jiménez Laredo, J. I. Hidalgo and K. O. Babaagba, editors, *Applications of Evolutionary Computation*, pages 738–754, Cham. Springer International Publishing, 2022.
- [26] A. Friedman. Stochastic Differential Equations and Applications. In *Stochastic differential equations*. volume **77**, C.I.M.E. Summer Sch. Pages 75–148. Springer, Heidelberg, 2010.
- [27] A. Garbuno-Inigo, F. Hoffmann, W. Li and A. M. Stuart. Interacting Langevin diffusions: gradient structure and ensemble Kalman sampler. *SIAM J. Appl. Dyn. Syst.*, **19**(1):412–441, 2020.
- [28] A. Garbuno-Inigo, N. Nüsken and S. Reich. Affine invariant interacting Langevin dynamics for Bayesian inference. *SIAM J. Appl. Dyn. Syst.*, **19**(3):1633–1658, 2020.
- [29] D. Gilbarg and N. S. Trudinger. *Elliptic Partial Differential Equations of Second Order*. Classics in Mathematics. Springer-Verlag, Berlin, 2001. Reprint of the 1998 edition.

- [30] S. Grassi and L. Pareschi. From particle swarm optimization to consensus based optimization: stochastic modeling and mean-field limit. *Math. Models Methods Appl. Sci.*, **31**(8):1625–1657, 2021.
- [31] S.-Y. Ha, S. Jin and D. Kim. Convergence of a first-order consensus-based global optimization algorithm. *Math. Models Methods Appl. Sci.*, **30**(12):2417–2444, 2020.
- [32] S.-Y. Ha, M. Kang, D. Kim, J. Kim and I. Yang. Convergence and error estimates for time-discrete consensus-based optimization algorithms. *Numerische Mathematik*:1–28, 2020.
- [33] S.-Y. Ha, M. Kang, D. Kim, J. Kim and I. Yang. Stochastic consensus dynamics for nonconvex optimization on the Stiefel manifold: mean-field limit and convergence. *Math. Models Methods Appl. Sci.*, **32**(3):533–617, 2022.
- [34] D. J. Higham, X. Mao and A. M. Stuart. Strong convergence of Euler-type methods for nonlinear stochastic differential equations. *SIAM J. Numer. Anal.*, **40**(3):1041–1063, 2002.
- [35] H. Huang and J. Qiu. On the mean-field limit for the consensus-based optimization. *Math. Methods Appl. Sci.*, **45**(12):7814–7831, 2022.
- [36] S. Jin, L. Li and J.-G. Liu. Random batch methods (RBM) for interacting particle systems. *J. Comput. Phys.*, **400**:108877, 30, 2020.
- [37] D. Kalise, A. Sharma and M. V. Tretyakov. Consensus-based optimization via jump-diffusion stochastic differential equations. *Math. Models Methods Appl. Sci.*, **33**(2):289–339, 2023.
- [38] J. Kennedy and R. Eberhart. Particle swarm optimization. In *Proceedings of ICNN'95-international conference on neural networks*, volume **4**, pages 1942–1948. IEEE, 1995.
- [39] R. Khasminskii. *Stochastic Stability of Differential Equations*, volume **66** of *Stochastic Modelling and Applied Probability*. Springer, Heidelberg, second edition, 2012. With contributions by G. N. Milstein and M. B. Nevelson.
- [40] F. Kittaneh. On Lipschitz functions of normal operators. *Proc. Amer. Math. Soc.*, **94**(3):416–418, 1985.
- [41] K. Klamroth, M. Stiglmayr and C. Totzeck. Consensus-based optimization for multi-objective problems: a multi-swarm approach. *arXiv preprint*, **2211.15737**, 2022.
- [42] D. Ko, S. Y. Ha, S. Jin and D. Kim. Convergence analysis of the discrete consensus-based optimization algorithm with random batch interactions and heterogeneous noises. *Math. Models Methods Appl. Sci.*, **32**(06):1071–1107, 2022.
- [43] X. Mao. *Stochastic Differential Equations and Applications*. Horwood Publishing Limited, Chichester, second edition, 2008.
- [44] M. Mitzenmacher and E. Upfal. *Probability and computing*. Cambridge University Press, Cambridge, second edition, 2017, pages xx+467. Randomization and probabilistic techniques in algorithms and data analysis.
- [45] B. Øksendal. *Stochastic Differential Equations*. Universitext. Springer-Verlag, Berlin, sixth edition, 2003. An introduction with applications.
- [46] R. Pinnau, C. Totzeck, O. Tse and S. Martin. A consensus-based model for global optimization and its mean-field limit. *Math. Models Methods Appl. Sci.*, **27**(1):183–204, 2017.
- [47] A.-S. Sznitman. Topics in propagation of chaos. In *École d'Été de Probabilités de Saint-Flour XIX—1989*. volume **1464**, Lecture Notes in Math. Pages 165–251. Springer, Berlin, 1991.
- [48] C. Totzeck. Trends in consensus-based optimization. In *Active particles. Vol. 3. Advances in theory, models, and applications*, Model. Simul. Sci. Eng. Technol. Pages 201–226. Birkhäuser/Springer, Cham, 2022.
- [49] C. Totzeck and M.-T. Wolfram. Consensus-based global optimization with personal best. *Math. Biosci. Eng.*, **17**(5):6026–6044, 2020.
- [50] U. Vaes. Sharp propagation of chaos for the ensemble Langevin sampler. *arXiv preprint*, **2404.06456**, 2024.
- [51] J. L. van Hemmen and T. Ando. An inequality for trace ideals. *Comm. Math. Phys.*, **76**(2):143–148, 1980.
- [52] C. Villani. *Optimal Transport*, volume **338** of *Grundlehren der mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 2009.
- [53] C. Villani. *Topics in Optimal Transportation*, volume **58** of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2003.